Research Institute
Social Cohesion

RDC

SOEP

SCP Documentation

German Social Cohesion Panel

SCP 2021-22 W1-2 Codebook HGEN: Houshold-Related Status and Generated Variables (English)

Research Institute Social Cohesion

SOEP The Socio-Economic Panel

German Social Cohesion Panel

Established in 2021, the German Social Cohesion Panel (SCP) is a wide-ranging representative longitudinal study of private households in Germany, carried out in collaboration of the Research Institute Social Cohesion (RISC) and the German Socio-Economic Panel (SOEP).

The aim of the SCP Documentation is to thoroughly document the survey's data collection and data processing.

Recommended Citation

Groh-Samberg, O., Axenfeld, J. B., Gerlitz, J.-Y., Cornesse, C., Kroh, M., Lengfeld, H., Liebig, S., Minkus, L., Reinecke, J., Teichler, N., Traunmüller, R., & Zinn, S. (2026). SCP 2021-22 W1-2 - Codebook HGEN: Household-Related Status and Generated Variables (English). *German Social Cohesion Panel 2021-2022 - Wave 1-2*. Bremen and Berlin: RDC-RISC/SOCIUM, SOEP/DIW Berlin. doi:10.60532/scp.2021-22.w1-2.v1

- **Authors:** Olaf Groh-Samberg, Julian B. Axenfeld, Jean-Yves Gerlitz, Carina Cornesse, Martin Kroh, Holger Lengfeld, Stefan Liebig, Lara Minkus, Jost Reinecke, Nils Teichler, Richard Traunmüller, Sabine Zinn

- **Contributors:** Cosima Adams, Anton Bochert, Martin Gerike, Josefine Kuhrmeier, Anna-Tabea Müller, Eric Nissen, Sebastian Rueda-Uribe, Rainer Siegers, Hans Walter Steinhauer, Knut Wenzig, Julia Witton (Project Members), infas (Data Collector)

- **Publisher:** RDC-RISC
  SOCIUM, University of Bremen
  P.O. Box 330 440
  28334 Bremen
  Germany

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

- **DOI:** 10.60532/scp.2021-22.w1-2.v1

- **Website:** www.fgz-risc-data.de

  www.diw.de

SCP Documentation

German Social Cohesion Panel

SCP 2021-22 W1-2 Codebook HGEN: Household-Related Status and Generated Variables (English)

Contents

1 General Information

The HGEN dataset contains user-friendly data on the level of the households. The data were generated using several sources, most importantly the H dataset. Each household (HID) that responded to the household questionnaire in a particular survey wave (WAVE) has one row in the dataset. Households whose anchor person did not take part in the survey wave or broke off the survey before the household questionnaire began are not included in the data set. Variables that were generated based on SOEP logic but with deviating measurement are marked with the suffix "_scp".
In some places in the documentation and in the data, year numbers are used, for example, for the names of variables and of the questionnaire instrument. These year numbers are always based on the field start of the data collection of the corresponding survey wave.
Survey variables might be missing, that is, lacking a valid code or value, for different reasons. In the SCP, 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, not know an answer, report invalid or implausible values, or break off the interview. In the second case, the interview design may exclude respondents with certain characteristics from some questions (e.g., owner-occupiers will not be asked about the amount of rent they pay). The following codes are used:
[-1] No answer
[-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
[-9] Don't want to answer [only CAWI]
[-10] Break-off [only CAWI]
[-1] No answer – Respondent declines to answer the question. Is assigned if the question is not answered in the paper questionnaire and if "I cannot answer that" is ticked in the online questionnaire. Note that the paper questionnaire does not distinguish between a refusal to answer and a true "don't know" except for a few questions where "don't know" is an additional answer option.
[-2] Does not apply – 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".
[-3] Implausible value – Sometimes invalid answers occur when respondents fill out a paper questionnaire (e.g., working hours over 168 per week, unreadable answers). In such cases the variable is recoded "-3".
[-4] Inadmissible multiple response – Some questions contain multiple answer options and respondents are asked to pick one answer. In the paper questionnaire 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 the code "-4" is assigned to this variable.
[-5] Not included in this version of the questionnaire – Respondents receive different questionnaire versions based on their respondent status (anchor person, household member, or

new household member) and their mode choice (paper or online questionnaire). Some questions are not part of one or more of these different questionnaire versions. For respondents who did not receive a question due to their questionnaire version, the variables are coded "-5".
[-6] Version of questionnaire with modified filtering – Currently not used.
[-7] Only available in less restricted edition – currently not used
[-8] Question not part of the survey program this year – Is assigned if the question is not asked in this particular survey year.
[-9] Don't want to answer [only CAWI] – Only applicable to respondents who answered the questionnaire online. Is assigned if "I do not want to answer that" is ticked in the online questionnaire.
[-10] Break-off [only CAWI] – Only applicable to respondents who answered the questionnaire online. Is assigned if respondents were not asked this question because they broke off the online questionnaire before completing it.

2 Identifiers

hid – Current Household ID

21100003		3
21100009		3
21100010		1
21100012		1
21100016		3
21100020		3
21100021		3
21100032		1
21100037		1
21100039		3
21100044		1
21100045		1
21100049		3
21100050		1
21100058		1
...	(13029 rows omitted)	25800
21137972		1
21137973		1
21137976		1
21137978		2
21137979		3
21137985		2
21137987		1
21137991		3
21138000		1
22103378		1
22103896		1
22115150		1
22117540		1
22119085		1

22125622	1

This identifier groups all individuals into their respective households at the time of the most recent wave (i.e. a person's HID can change over time, for example if an adult child moves out of their parents' home and starts their own household).

cid – Original Household ID

21100003		3
21100009		3
21100010		1
21100012		1
21100016		3
21100020		3
21100021		3
21100032		1
21100037		1
21100039		3
21100044		1
21100045		1
21100049		3
21100050		1
21100058		1
...	(13023 rows omitted)	25795
21137960		1
21137961		1
21137963		3
21137964		3
21137967		2
21137971		1
21137972		1
21137973		1
21137976		1
21137978		2
21137979		3
21137985		2
21137987		1
21137991		3
21138000		1

This identifier groups individuals into their original households at the start of the panel. That means that a person's CID is time-constant and will always relate them back to the household they initially belonged to, even if they moved out since.

3 Survey Context

wave – Survey Wave

1	[1] Wave 1, part 1 (2021/22)	13053
2	[2] Wave 1, part 2 (2021/22)	6669
3	[3] Wave 2 (2022/23)	6128

This variable identifies the (partial) wave in which the data collection took place.

hghmonth – HH-Related Survey Month

? Interview: date *(from: scp/w2/interviewdatum:interviewdatum[])*

1	[1] January	2394
2	[2] February	1270
3	[3] March	684
4	[4] April	404
5	[5] May	723
6	[6] June	178
7	[7] July	48
8	[8] August	0
9	[9] September	11495
10	[10] October	5742
11	[11] November	1145
12	[12] December	1767

This variable identifies the month in which an anchor person participated in a data collection wave.

iyear – HH-Related Survey Year

? Interview: date *(from: scp/w2/interviewdatum:interviewdatum[])*

2021	14024
2022	11823
2023	3

This variable identifies the year in which an anchor person participated in a data collection wave. IYEAR can vary within a data collection wave if data collection includes the turn of a year.

hghmode – Survey Mode

? Survey Mode *(from: scp/w2/pgmode:pgmode[])*

210	[210] Paper questionnaire, postal (PAPI)	13891
300	[300] Online questionnaire (CAWI)	11959

The survey mode is determined by the technology respondents use to fill out a questionnaire. To prevent duplications, if respondents submit a questionnaire via multiple modes, priority is given to the data they provided via the web.

4 Regional Information

bula_scp – Federal State

5	[5] North Rhine-Westphalia	2797
6	[6] Hesse	1024
8	[8] Baden-Württemberg	2033
9	[9] Bavaria	2579
11	[11] Berlin	878
12	[12] Brandenburg	1283
13	[13] Mecklenburg-Western Pomerania	1011
14	[14] Saxony	1878
15	[15] Saxony-Anhalt	888
16	[16] Thuringia	919
17	[17] Bremen/Lower Saxony	1849
18	[18] Hamburg/Schleswig-Holstein	946
19	[19] Saarland/Rhineland-Palatinate	1091
-1	[-1] No answer	5
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	6669
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

The information on the federal state is not part of the survey data but was taken from the sampling frame (i.e. population register data). German federal states with few participating households were combined with their neighbouring state. Bremen and Lower Saxony form one category, Hamburg and Schleswig-Holstein form another and Saarland was combined with Rhineland-Palatinate.

hgnuts1_scp – NUTS Systematic 1 (Federal State)

1	[1] [DE100] Baden-Württemberg	2033
2	[2] [DE200] Bavaria	2579
3	[3] [DE300] Berlin	878
4	[4] [DE400] Brandenburg	1283
7	[7] [DE700] Hesse	1024
8	[8] [DE800] Mecklenburg-Western Pomerania	1011
10	[10] [DEA00] North Rhine-Westphalia	2797
13	[13] [DED00] Saxony	1878
14	[14] [DEE00] Saxony-Anhalt	888
16	[16] [DEG00] Thuringia	919
17	[17] [DE500] HB/[DE900] NI	1849
18	[18] [DE600] HH/[DEF00] SH	946
19	[19] [DEB00] RP/[DEC00] SL	1091

-1	[-1] No answer	5
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	6669
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

NUTS ("Nomenclature of Statistical Territorial Units") is a hierarchical system for dividing up the economic territory of the European Union. It was introduced by Eurostat more than 30 years ago in order to provide a single uniform breakdown of territorial units for the production of regional statistics. NUTS 1 especially contains the major socio-economic regions for analyzing regional community problems. It currently divides the European Union into 97 regions. In Germany, these are equivalent to the 16 German Federal States. German Federal States with few participating households were combined with their neighbouring state. Bremen and Lower Saxony form one category, Hamburg and Schleswig-Holstein form another and Saarland was combined with Rhineland-Palatinate.

ostwest – Eastern vs. Western Germany

1	[1] West	12319
2	[2] East (incl. Berlin)	6857
-1	[-1] No answer	5
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	6669
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

The information on the part of the country is not part of the survey data but was taken from the sampling frame (i.e. population register data).

5 Household Context

hghhsize – HH size (generated)

? Survey Mode *(from: scp/w2/pgmode:pgmode[])*
? Interview: date *(from: scp/w2/interviewdatum:interviewdatum[])*
? When were you born? // Month [...] *(from: scp/w2/Q45:pbirthmnth[])*
? When were you born? // Year [...] *(from: scp/w2/Q45:pbirthyear[])*
? Are you living with a partner in the household? *(from: scp/w2/Q47:pcohab[])*

? How many people live permanently in your household, children, and _yourself included_? // Number: [...] people in household *(from: scp/w2/Q50:phhsize[])*
? How many persons who live in your household all the time (_including yourself_), ... // Number: [...] people *(from: scp/w2/Q51:phhage01[])*
? How many persons who live in your household all the time (_including yourself_), ... // Number: [...] people *(from: scp/w2/Q51:phhage02[])*
? How many persons who live in your household all the time (_including yourself_), ... // There are no children under the age of 14 in my household *(from: scp/w2/Q51:phhage03[])*
? What is this [number of loop iteration] person's first name? // First name of the person: [...] *(from: scp/w2/Q123a1:hpnameX[])*
? What month and year was [first name number of loop iteration] born? Date of birth: // Month of birth: [...] *(from: scp/w2/Q123a4:hpbmnthX[])*
? What month and year was [first name number of loop iteration] born? Date of birth: // Year of birth: [...] *(from: scp/w2/Q123a4:hpbyearX[])*

1		4114
2		8494
3		3012
4		2368
5		683
6		175
7		52
8		13
9		8
10		13
11		3
12		2
13		1
16		1
17		1
18		1
22		1
30		1
-1	[-1] No answer	238
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	6669
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

The generated variables HGHHSIZE, HGHHAGE01, HGHHAGE02 and HGHHAGE03 were generated simultaneously, with a focus on harmonization. The first step was to aggregate the AP information from the HH matrix and check those and the direct information on HH size and age groups for consistency. If there were large inconsistencies in the information of the APs and more than one interview was available, the direct information on

HH size and age groups of the HMs was checked for consistency and used. If both the APs' and HMs' data were highly inconsistent or missing, the HH size of HHs with more than one invited person was set to the number of invited persons (and all to age group 18+), while HHs in which only the AP was invited were set to missing. Source variables: PHHSIZE, PHHAGE01, PHHAGE02, PHHAGE03, HPNAME01-HPNAME12, HPBMNTH01-HPBMNTH12, HPBYEAR01-HPBYEAR12, PCOHAB and PID. Inconsistencies are shown in the variable HGHHSIZE_FLAG. The variables wer filled step by step:

1. AP: include fully consistent cases
2. AP: include slighty inconsistent cases if HH sizes (matrix & response) & generated age groups are consistent
3. AP: include cases if HH size response is missing, but matrix information is consistent
4. AP: include cases if HH size response is smaller than the generated one, but whose matrix information is consistent
5. AP: include cases if HH size & age group responses is are consistent
6. AP: include information with partial missings or inconsistencies if only 1 interview & the reported HH size was greater than 1
7. HMs: include fully consistent cases (intra-person & inter-person)
8. HMs: include cases if the (intra-person & inter-person) information of at least 2 cases is consistent
9. HMs: include cases if the (intra-person) information is consistent, starting with the oldest
10. PBRUTTO: assign no. of invited persons if more than 1 invited person in HH
11. PBRUTTO: assign -1 if 1 invited person in HH

hghhage01 – Number HH members 0-13 years (generated)

? Survey Mode *(from: scp/w2/pgmode:pgmode[])*
? Interview: date *(from: scp/w2/interviewdatum:interviewdatum[])*
? When were you born? // Month [...] *(from: scp/w2/Q45:pbirthmnth[])*
? When were you born? // Year [...] *(from: scp/w2/Q45:pbirthyear[])*
? Are you living with a partner in the household? *(from: scp/w2/Q47:pcohab[])*
? How many people live permanently in your household, children, and _yourself included_? // Number: [...] people in household *(from: scp/w2/Q50:phhsize[])*
? How many persons who live in your household all the time (_including yourself_), ... // Number: [...] people *(from: scp/w2/Q51:phhage01[])*
? How many persons who live in your household all the time (_including yourself_), ... // Number: [...] people *(from: scp/w2/Q51:phhage02[])*
? How many persons who live in your household all the time (_including yourself_), ... // There are no children under the age of 14 in my household *(from: scp/w2/Q51:phhage03[])*
? What is this [number of loop iteration] person's first name? // First name of the person: [...] *(from: scp/w2/Q123a1:hpnameX[])*
? What month and year was [first name number of loop iteration] born?

Date of birth: // Month of birth: [...] *(from: scp/w2/Q123a4:hpbmnthX[])*
? What month and year was [first name number of loop iteration] born?
Date of birth: // Year of birth: [...] *(from: scp/w2/Q123a4:hpbyearX[])*

0		15541
1		1822
2		1280
3		260
4		28
5		6
6		3
8		1
11		1
13		1
-1	[-1] No answer	238
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	6669
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

See description of HGHHSIZE.

hghhage02 – Number HH members 14-17 years (generated)

? Survey Mode *(from: scp/w2/pgmode:pgmode[])*
? Interview: date *(from: scp/w2/interviewdatum:interviewdatum[])*
? When were you born? // Month [...] *(from: scp/w2/Q45:pbirthmnth[])*
? When were you born? // Year [...] *(from: scp/w2/Q45:pbirthyear[])*
? Are you living with a partner in the household? *(from: scp/w2/Q47:pcohab[])*
? How many people live permanently in your household, children, and _yourself included_? // Number: [...] people in household *(from: scp/w2/Q50:phhsize[])*
? How many persons who live in your household all the time (_including yourself_), ... // Number: [...] people *(from: scp/w2/Q51:phhage01[])*
? How many persons who live in your household all the time (_including yourself_), ... // Number: [...] people *(from: scp/w2/Q51:phhage02[])*
? How many persons who live in your household all the time (_including yourself_), ... // There are no children under the age of 14 in my household *(from: scp/w2/Q51:phhage03[])*
? What is this [number of loop iteration] person's first name? // First name of the person: [...] *(from: scp/w2/Q123a1:hpnameX[])*
? What month and year was [first name number of loop iteration] born?
Date of birth: // Month of birth: [...] *(from: scp/w2/Q123a4:hpbmnthX[])*
? What month and year was [first name number of loop iteration] born?
Date of birth: // Year of birth: [...] *(from: scp/w2/Q123a4:hpbyearX[])*

0		17481
1		1247
2		197
3		12
4		5
7		1
-1	[-1] No answer	238
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	6669
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

See description of HGHHSIZE.

hghhage03 – Number HH members 18+ years (generated)

? Survey Mode *(from: scp/w2/pgmode:pgmode[])*
? Interview: date *(from: scp/w2/interviewdatum:interviewdatum[])*
? When were you born? // Month [...] *(from: scp/w2/Q45:pbirthmnth[])*
? When were you born? // Year [...] *(from: scp/w2/Q45:pbirthyear[])*
? Are you living with a partner in the household? *(from: scp/w2/Q47:pcohab[])*
? How many people live permanently in your household, children, and _yourself included_? // Number: [...] people in household *(from: scp/w2/Q50:phhsize[])*
? How many persons who live in your household all the time (_including yourself_), ... // Number: [...] people *(from: scp/w2/Q51:phhage01[])*
? How many persons who live in your household all the time (_including yourself_), ... // Number: [...] people *(from: scp/w2/Q51:phhage02[])*
? How many persons who live in your household all the time (_including yourself_), ... // There are no children under the age of 14 in my household *(from: scp/w2/Q51:phhage03[])*
? What is this [number of loop iteration] person's first name? // First name of the person: [...] *(from: scp/w2/Q123a1:hpnameX[])*
? What month and year was [first name number of loop iteration] born? Date of birth: // Month of birth: [...] *(from: scp/w2/Q123a4:hpbmnthX[])*
? What month and year was [first name number of loop iteration] born? Date of birth: // Year of birth: [...] *(from: scp/w2/Q123a4:hpbyearX[])*

1	4516
2	11356
3	2118
4	728
5	138
6	45
7	14

8		4
9		7
10		9
11		3
16		1
17		2
18		1
22		1
-1	[-1] No answer	238
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	6669
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

See description of HGHHSIZE.

hghhsize_flag – Flag HH size and age groups

? Survey Mode *(from: scp/w2/pgmode:pgmode[])*
? Interview: date *(from: scp/w2/interviewdatum:interviewdatum[])*
? When were you born? // Month [...] *(from: scp/w2/Q45:pbirthmnth[])*
? When were you born? // Year [...] *(from: scp/w2/Q45:pbirthyear[])*
? Are you living with a partner in the household? *(from: scp/w2/Q47:pcohab[])*
? How many people live permanently in your household, children, and _yourself included_? // Number: [...] people in household *(from: scp/w2/Q50:phhsize[])*
? How many persons who live in your household all the time (_including yourself_), ... // Number: [...] people *(from: scp/w2/Q51:phhage01[])*
? How many persons who live in your household all the time (_including yourself_), ... // Number: [...] people *(from: scp/w2/Q51:phhage02[])*
? How many persons who live in your household all the time (_including yourself_), ... // There are no children under the age of 14 in my household *(from: scp/w2/Q51:phhage03[])*
? What is this [number of loop iteration] person's first name? // First name of the person: [...] *(from: scp/w2/Q123a1:hpnameX[])*
? What month and year was [first name number of loop iteration] born? Date of birth: // Month of birth: [...] *(from: scp/w2/Q123a4:hpbmnthX[])*
? What month and year was [first name number of loop iteration] born? Date of birth: // Year of birth: [...] *(from: scp/w2/Q123a4:hpbyearX[])*

[-8] Question not part of the survey program this year	6669
0000000000000	14105
0000000000011	527
0000000000100	1
0000000000101	31

0000000000110		135
0000000000111		7
0001000000011		3
0001000000110		4
0001000000111		2
0001000100111		196
0001001000011		313
0001001000111		12
0001010000110		56
0001010000111		204
...	(135 rows omitted)	3491
1101010000110-1001000000		1
1101010000111		1
1101010000111-0000000000		1
1101010000111-1000010000		1
1110000000000		45
1110000000001		20
1110000000011		4
1110000000100		1
1110000000101		4
1110000000111		2
1111000010111		3
1111000010111-0000000000		7
1111000010111-0000001100		1
1111000100111-0000000000		1
1111010000111-0000000000		2

This flag variable indicates the reliability of the information of the generated variables HGHH-SIZE, HGHHAGE01, HGHHAGE02 and HGHHAGE03.
Flag = 0000000000000-0000000000: AP; HH sizes matrix, aggregated response & no. of interviews - full consistency
Flag = 1xxxxxxxxxxxx-xxxxxxxxxx: AP; HH size of matrix & no. of interviews - more interviews than HM in HH matrix
Flag = x1xxxxxxxxxxx-xxxxxxxxxx: AP; HH sizes of matrix & aggregated response - mismatch or missing
Flag = xx1xxxxxxxxx-xxxxxxxxxx: AP; HH size matrix & age groups matrix - mismatch or missing
Flag = xxx1xxxxxxxx-xxxxxxxxxx: AP; HH size response & age groups response - mismatch or missing
Flag = xxxx1xxxxxxx-xxxxxxxxxx: AP; HH size response & partnership info - mismatch
Flag = xxxxx1xxxxxx-xxxxxxxxxx: AP; HH size & age group 18+ responses available, age group 01-13 response missing
Flag = xxxxxx1xxxxx-xxxxxxxxxx: AP; HH size & age group 01-13 responses available, age group 18+ response missing
Flag = xxxxxxx1xxxx-xxxxxxxxxx: AP; HH size response available, age groups 01-13 & 18+ responses missing
Flag = xxxxxxxx1xxx-xxxxxxxxxx: AP; HH size response missing
Flag = xxxxxxxxx1xx-xxxxxxxxxx: AP; HH size reponse & no. of interviews - more interviews than HM in HH response

Flag = xxxxxxxxxx1xx-xxxxxxxxxx: AP; age group 0-13 matrix & response - mismatch or missing
Flag = xxxxxxxxxx1x-xxxxxxxxxx: AP; age group 14-17 matrix & response - mismatch or missing
Flag = xxxxxxxxxx1-xxxxxxxxxx: AP; age group 18+ matrix & response - mismatch or missing
Flag = xxxxxxxxxxxx-0000000000: HMs; HH sizes matrix, aggregated response & no. of interviews - full consistency
Flag = xxxxxxxxxxxx-1xxxxxxxxx: HMs; HH size response & age groups responses - mismatch or missing
Flag = xxxxxxxxxxxx-x1xxxxxxx: HMs; HH size response & partnership info - mismatch
Flag = xxxxxxxxxxxx-xx1xxxxxx: HMs; HH size & age group 18+ responses available, age group 01-13 response missing
Flag = xxxxxxxxxxxx-xxx1xxxxx: HMs; HH size & age group 01-13 responses available, age group 18+ response missing
Flag = xxxxxxxxxxxx-xxxx1xxxx: HMs; HH size responses available, age groups 01-13 & 18+ response missing
Flag = xxxxxxxxxxxx-xxxxx1xxx: HMs; HH size response missing
Flag = xxxxxxxxxxxx-xxxxx1xxx: HMs; HH size response across multiple HMs - mismatch or missing
Flag = xxxxxxxxxxxx-xxxxxx1xx: HMs; age group 18+ response across multiple HMs - mismatch or missing
Flag = xxxxxxxxxxxx-xxxxxxx1x: HMs; age group 1-13 response across multiple HMs - mismatch or missing
Flag = xxxxxxxxxxxx-xxxxxxxx1: HMs; HH-size response & no. of interviews - more interviews
0 = issue did not occur, 1 = issue occured, x = either/or

6 Dwelling

hgsize – Size of Housing Unit in Sq M

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		3
14		1
15		7
16		6
17		5
18		6
19		2
20		19
21		6
22		5
23		14
24		12
...	(242 rows omitted)	11902

573		1
580		1
600		3
800		1
1000		1
-1	[-1] No answer	564
-2	[-2] Does not apply	0
-3	[-3] Implausible value	2
-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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	91
-10	[-10] Break-off [only CAWI]	138

This variable contains the size of the housing unit as reported by the anchor person without imputations for item-nonresponse.

hgi1size – 1. Imputed Size of Housing Unit in Sq M [1/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		3
14		1
15		7
16		7
17		5
18		6
19		2
20		20
21		6
22		7
23		15
24		13
...	(241 rows omitted)	12689
573		1
580		2
600		4
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 1st of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006).
Multiple imputation is a technique to deal with missing data in which missing values are replaced by multiple imputed values that are plausible based on a statistical imputation model. These imputed values serve not as an exact estimate of the missing value, but to recreate both relationships between variables and the variance and uncertainty in the data. To analyse the multiply imputed data, the statistical analysis of interest needs to be performed on all imputed versions of the dataset (in this case, 15) separately, and the estimates from these separate analyses are combined into single estimates according to specific combining rules thereafter. Conventional statistics software often includes dedicated solutions for dealing with multiply imputed data, such as the mice package in R (R Core Team, 2023; van Buuren & Groothuis-Oudshoorn, 2011) or the built-in command mi estimate in Stata (StataCorp, 2023).
The imputations in HGI1SIZE are intended primarily to ease univariate analyses on the whole sample, both cross-sectional and longitudinal, and are based on the assumption that the data are missing at random (MAR). Because it is impossible to cover all relationships in the data with the imputation model, estimates in bivariate and multivariate analyses as well as subgroup analyses may be biased to some extent on a case-by-case basis, as the congeniality principle (Meng, 1994) may be violated. However, all imputed variables in the SCP (individual gross labour income, individual net labour income, household net income, and size of housing unit) were imputed together in an iterative fashion through multiple imputation by chained equations, which means they can be analyzed together in a statistical model if needed. For more complex analyses of the data, users may consider applying multiple imputation by themselves using an imputation model tailored to their specific substantive model of interest.
This variable was imputed using type-1 predictive mean matching as implemented in the mice package in R. Simply put, this method entails (1) estimating predictive means of the variable to be imputed using a regression model and (2) using these predictive means to match missing with observed values, which serve as donors for the imputations. To deal with the highly skewed distribution of the imputed variable, a log transformation was applied to the variable before imputation, with the imputed variable being transformed back to the initial scale after imputation.
Predictor variables in the imputation model cover a broad selection of variables from the survey and additional microgeographic data. Variables from other waves were included as additional predictor variables, with the data being reshaped to the wide format. Predictors on the individual level were aggregated to the household level by calculating their household means and standard deviations and including them as predictors instead of the original variables. Predictors for the imputation model were selected through a threshold of r=0.2 in terms of their correlation to the imputed variable and, subsequently, by lasso regression with cross-validated lambda parameter. For further dimensionality reduction of the predictor space to a maximum of 40 components, partial least squares regression was applied to the selected predictor variables.

Meng, X. L. (1994). Multiple-imputation inferences with uncongenial sources of input. Statistical Science, 9(4), 538-558.
R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/
Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. Wiley.
StataCorp (2023). Stata 18 Multiple-Imputation Reference Manual. Stata Press.
van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. Journal of Statistical Computation and Simulation, 76(12), 1049–1064.
van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. Journal of Statistical Software, 45(3), 1-67.

hgi2size – 2. Imputed Size of Housing Unit in Sq M [2/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		3
14		2
15		8
16		6
17		5
18		6
19		4
20		21
21		6
22		6
23		15
24		12
...	(241 rows omitted)	12688
573		1
580		1
600		3
800		2
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 2nd of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi3size – 3. Imputed Size of Housing Unit in Sq M [3/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		4
14		1
15		7
16		6
17		5
18		6
19		2
20		19
21		7
22		6
23		16
24		14
...	(241 rows omitted)	12688
573		1
580		1
600		5
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 3rd of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi4size – 4. Imputed Size of Housing Unit in Sq M [4/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		3
14		1
15		8
16		7
17		5
18		7
19		2
20		19
21		8
22		6
23		15
24		12
...	(241 rows omitted)	12687
573		1
580		1
600		6
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 4th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi5size – 5. Imputed Size of Housing Unit in Sq M [5/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4

13		3
14		2
15		7
16		6
17		5
18		6
19		2
20		20
21		6
22		5
23		16
24		13
...	(241 rows omitted)	12691
573		1
580		1
600		4
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 5th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi6size – 6. Imputed Size of Housing Unit in Sq M [6/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		6
13		3
14		2
15		7
16		6
17		5
18		6
19		2
20		20

21		6
22		6
23		17
24		13
...	(241 rows omitted)	12687
573		1
580		1
600		4
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 6th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi7size – 7. Imputed Size of Housing Unit in Sq M [7/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		4
14		1
15		7
16		6
17		5
18		8
19		2
20		21
21		6
22		7
23		16
24		13
...	(241 rows omitted)	12685
573		1
580		1
600		5

800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 7th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi8size – 8. Imputed Size of Housing Unit in Sq M [8/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		3
14		1
15		7
16		6
17		5
18		7
19		2
20		23
21		7
22		6
23		14
24		13
...	(241 rows omitted)	12688
573		1
580		1
600		4
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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

-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 8th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi9size – 9. Imputed Size of Housing Unit in Sq M [9/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		4
14		1
15		9
16		7
17		5
18		7
19		2
20		21
21		6
22		7
23		14
24		12
...	(241 rows omitted)	12686
573		1
580		2
600		4
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 9th of 15 imputations as generated by multiple imputation by chained

equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi10size – 10. Imputed Size of Housing Unit in Sq M [10/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		2
10		2
12		5
13		3
14		1
15		7
16		8
17		5
18		7
19		4
20		23
21		6
22		5
23		14
24		13
...	(241 rows omitted)	12685
573		1
580		1
600		3
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 10th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi11size – 11. Imputed Size of Housing Unit in Sq M [11/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		3
14		1
15		7
16		6
17		5
18		6
19		3
20		21
21		6
22		5
23		15
24		14
...	(241 rows omitted)	12688
573		2
580		1
600		5
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 11th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi12size – 12. Imputed Size of Housing Unit in Sq M [12/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		3
14		1
15		9
16		6
17		5

18		6
19		2
20		20
21		6
22		8
23		15
24		16
...	(241 rows omitted)	12685
573		1
580		1
600		3
800		1
1000		2
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 12th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi13size – 13. Imputed Size of Housing Unit in Sq M [13/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		4
14		1
15		8
16		7
17		5
18		6
19		2
20		19
21		6
22		5
23		15
24		13
...	(241 rows omitted)	12692

573		1
580		1
600		3
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 13th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi14size – 14. Imputed Size of Housing Unit in Sq M [14/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		1
10		2
12		4
13		3
14		2
15		7
16		6
17		5
18		6
19		2
20		20
21		7
22		5
23		17
24		14
...	(241 rows omitted)	12688
573		1
580		1
600		4
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 14th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgi15size – 15. Imputed Size of Housing Unit in Sq M [15/15]

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

7		2
10		2
12		4
13		3
14		1
15		7
16		6
17		5
18		6
19		2
20		22
21		7
22		5
23		16
24		13
...	(241 rows omitted)	12689
573		1
580		1
600		3
800		1
1000		1
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the size of the housing unit in square metres (HGSIZE) with missing values replaced by the 15th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1SIZE.

hgfsize – Imputation Flag of Housing Size

? How large is the total living space in this dwelling? // Living space [...] square metres *(from: scp/w2/Q128:hflatsize[])*

0	[0] Not imputed	25055
1	[1] Imputed	795
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable designates imputations of item-nonresponse in the variable HGSIZE (size of housing unit in square meters).

hgowner_scp – Tenant or Owner of Dwelling

? Is it a boarding house or similar accommodation? *(from: scp/w2/Q130:hdorm01[])*
? Are you the main tenant, subletter, or owner? *(from: scp/w2/Q132:htenancy[])*

1	[1] Owner	6297
2	[2] Main Tenant	5378
3	[3] Sub-tenant	770
4	[4] Tenant	0
5	[5] Home Occupant	13
-1	[-1] No answer	161
-2	[-2] Does not apply	0
-3	[-3] Implausible value	0
-4	[-4] Inadmissible multiple response	5
-5	[-5] Not included in this version of the questionnaire	0
-6	[-6] Version of questionnaire with modified filtering	0
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	13053
-9	[-9] Don't want to answer [only CAWI]	31
-10	[-10] Break-off [only CAWI]	142

The variable was generated from HTENANCY and HDORM01. For generation in wave 1 part 2, PHHTYP from wave 1 part 1 was also used; if there were missing values in HTENANCY

and HDORM01 and PHHTYP from wave 1 part 1 had the value 4 ("a retirement or nursing home or home for the disabled where you do not run an independent household, i.e. do not manage your own affairs"), the observations were coded as 5 ("Home Occupant").

7 Income

hghinc – Monthly HH Net Inc. (Euros)

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		410
1		13
2		1
3		2
5		1
8		1
22		2
30		1
50		2
52		1
100		4
125		1
140		1
141.02		1
150		2
...	(1977 rows omitted)	21505
400000		1
549497		1
999999		1
1000000		1
7100060		1
-1	[-1] No answer	2293
-2	[-2] Does not apply	142
-3	[-3] Implausible value	9
-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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	1012
-10	[-10] Break-off [only CAWI]	441

This variable contains the monthly net household income. This typically corresponds to the household net income as reported by the anchor person in HHHINCNET. In W1T1, the variable is generated also primarily from the anchor person's response, but is replaced by the respone of a household member from PHHINCNET if the anchor person does not report the income or the reported income is invalid.

hgi1hinc – 1. Imputed Monthly HH Net Inc. (Euros) [1/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		474
1		14
2		2
3		2
5		1
8		1
22		3
30		2
50		2
52		1
100		7
125		1
140		1
141.02		2
150		2
...	(1975 rows omitted)	25186
400000		1
549497		1
999999		1
1000000		1
7100060		3
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 1st of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006).
Multiple imputation is a technique to deal with missing data in which missing values are replaced by multiple imputed values that are plausible based on a statistical imputation model. These imputed values serve not as an exact estimate of the missing value, but to recreate both relationships between variables and the variance and uncertainty in the data. To analyse the multiply imputed data, the statistical analysis of interest needs to be performed on all imputed versions of the dataset (in this case, 15) separately, and the estimates from these separate analyses are combined into single estimates according to specific combining rules thereafter. Conventional statistics software often includes dedicated solutions for dealing with multiply imputed data, such as the mice package in R (R Core Team, 2023; van Buuren

& Groothuis-Oudshoorn, 2011) or the built-in command mi estimate in Stata (StataCorp, 2023).
The imputations in HGI1HINC are intended primarily to ease univariate analyses on the whole sample, both cross-sectional and longitudinal, and are based on the assumption that the data are missing at random (MAR). Because it is impossible to cover all relationships in the data with the imputation model, estimates in bivariate and multivariate analyses as well as subgroup analyses may be biased to some extent on a case-by-case basis, as the congeniality principle (Meng, 1994) may be violated. However, all imputed variables in the SCP (individual gross labour income, individual net labour income, household net income, and size of housing unit) were imputed together in an iterative fashion through multiple imputation by chained equations, which means they can be analyzed together in a statistical model if needed. For more complex analyses of the data, users may consider applying multiple imputation by themselves using an imputation model tailored to their specific substantive model of interest.
This variable was imputed using type-1 predictive mean matching as implemented in the mice package in R. Simply put, this method entails (1) estimating predictive means of the variable to be imputed using a regression model and (2) using these predictive means to match missing with observed values, which serve as donors for the imputations. To deal with the highly skewed distribution of the imputed variable, a log transformation was applied to the variable before imputation, with the imputed variable being transformed back to the initial scale after imputation.
Predictor variables in the imputation model cover a broad selection of variables from the survey and additional microgeographic data. Variables from other waves were included as additional predictor variables, with the data being reshaped to the wide format. Predictors on the individual level were aggregated to the household level by calculating their household means and standard deviations and including them as predictors instead of the original variables. Predictors for the imputation model were selected through a threshold of r=0.2 in terms of their correlation to the imputed variable and, subsequently, by lasso regression with cross-validated lambda parameter. For further dimensionality reduction of the predictor space to a maximum of 40 components, partial least squares regression was applied to the selected predictor variables.
Meng, X. L. (1994). Multiple-imputation inferences with uncongenial sources of input. Statistical Science, 9(4), 538-558.
R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/
Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. Wiley.
StataCorp (2023). Stata 18 Multiple-Imputation Reference Manual. Stata Press.
van Buuren, S., Brand, J. P., Groothuis-Oudshoorn, C. G., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. Journal of Statistical Computation and Simulation, 76(12), 1049–1064.
van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. Journal of Statistical Software, 45(3), 1-67.

hgi2hinc – 2. Imputed Monthly HH Net Inc. (Euros) [2/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0	492
1	15

2		1
3		2
5		1
8		1
22		2
30		1
50		2
52		1
100		8
125		1
140		1
141.02		1
150		2
...	(1975 rows omitted)	25170
400000		1
549497		1
999999		2
1000000		2
7100060		1
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 2nd of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi3hinc – 3. Imputed Monthly HH Net Inc. (Euros) [3/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0	569
1	13
2	2
3	2
5	1
8	1
22	2
30	1
50	2

52		1
100		5
125		1
140		1
141.02		1
150		2
...	(1975 rows omitted)	25099
400000		1
549497		1
999999		1
1000000		1
7100060		1
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 3rd of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi4hinc – 4. Imputed Monthly HH Net Inc. (Euros) [4/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		485
1		16
2		2
3		3
5		1
8		2
22		2
30		2
50		3
52		1
100		5
125		1
140		1
141.02		1
150		3
...	(1975 rows omitted)	25174

400000		1
549497		1
999999		2
1000000		1
7100060		1
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 4th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi5hinc – 5. Imputed Monthly HH Net Inc. (Euros) [5/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		481
1		14
2		1
3		2
5		1
8		2
22		3
30		2
50		3
52		1
100		8
125		2
140		1
141.02		1
150		4
...	(1975 rows omitted)	25175
400000		1
549497		2
999999		1
1000000		2
7100060		1
-1	[-1] No answer	0
-2	[-2] Does not apply	142

-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 5th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi6hinc – 6. Imputed Monthly HH Net Inc. (Euros) [6/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		480
1		18
2		1
3		2
5		1
8		1
22		3
30		1
50		2
52		1
100		7
125		1
140		1
141.02		3
150		3
...	(1975 rows omitted)	25177
400000		2
549497		1
999999		1
1000000		1
7100060		1
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0

-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 6th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi7hinc – 7. Imputed Monthly HH Net Inc. (Euros) [7/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		475
1		13
2		1
3		4
5		1
8		2
22		2
30		1
50		2
52		1
100		4
125		1
140		1
141.02		3
150		5
...	(1975 rows omitted)	25185
400000		2
549497		1
999999		1
1000000		1
7100060		2
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 7th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi8hinc – 8. Imputed Monthly HH Net Inc. (Euros) [8/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		547
1		16
2		2
3		2
5		1
8		1
22		2
30		2
50		2
52		1
100		6
125		1
140		1
141.02		1
150		3
...	(1975 rows omitted)	25115
400000		1
549497		1
999999		1
1000000		1
7100060		1
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 8th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi9hinc – 9. Imputed Monthly HH Net Inc. (Euros) [9/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		467

1		19
2		1
3		2
5		1
8		1
22		3
30		1
50		2
52		1
100		5
125		1
140		1
141.02		1
150		3
...	(1975 rows omitted)	25192
400000		2
549497		1
999999		1
1000000		1
7100060		2
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 9th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi10hinc – 10. Imputed Monthly HH Net Inc. (Euros) [10/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		455
1		18
2		1
3		2
5		1
8		1
22		4
30		3

50		2
52		1
100		6
125		1
140		1
141.02		1
150		2
...	(1975 rows omitted)	25202
400000		2
549497		1
999999		1
1000000		2
7100060		1
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 10th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi11hinc – 11. Imputed Monthly HH Net Inc. (Euros) [11/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		488
1		18
2		2
3		2
5		1
8		1
22		3
30		2
50		2
52		1
100		4
125		1
140		1
141.02		2
150		4

...	(1975 rows omitted)	25171
400000		1
549497		1
999999		1
1000000		1
7100060		1
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 11th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi12hinc – 12. Imputed Monthly HH Net Inc. (Euros) [12/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		455
1		15
2		2
3		2
5		1
8		1
22		2
30		1
50		3
52		1
100		8
125		1
140		1
141.02		2
150		4
...	(1975 rows omitted)	25204
400000		1
549497		1
999999		1
1000000		1
7100060		1
-1	[-1] No answer	0

-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 12th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi13hinc – 13. Imputed Monthly HH Net Inc. (Euros) [13/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		505
1		14
2		2
3		2
5		1
8		1
22		2
30		2
50		2
52		1
100		6
125		1
140		1
141.02		1
150		3
...	(1975 rows omitted)	25157
400000		1
549497		1
999999		2
1000000		2
7100060		1
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0

-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 13th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi14hinc – 14. Imputed Monthly HH Net Inc. (Euros) [14/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		455
1		19
2		1
3		2
5		1
8		1
22		3
30		2
50		2
52		1
100		6
125		1
140		1
141.02		2
150		4
...	(1975 rows omitted)	25199
400000		1
549497		1
999999		2
1000000		1
7100060		3
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 14th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgi15hinc – 15. Imputed Monthly HH Net Inc. (Euros) [15/15]

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0		478
1		14
2		1
3		2
5		1
8		1
22		2
30		1
50		2
52		1
100		5
125		1
140		1
141.02		2
150		2
...	(1975 rows omitted)	25189
400000		1
549497		1
999999		1
1000000		1
7100060		1
-1	[-1] No answer	0
-2	[-2] Does not apply	142
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable contains the net household net income (HGHINC) with missing values replaced by the 15th of 15 imputations as generated by multiple imputation by chained equations (Rubin, 1987; van Buuren et al., 2006). For more information on the imputations, see HGI1HINC.

hgfhinc – Imputation Flag of Monthly Net HH Inc.

? If you look at the total income of all of the members of your household: what is your monthly household income today? // Monthly household income: [...] Euros per month *(from: scp/w2/Q149:hhhincnet[])*

0	[0] Not imputed	22095

1	[1] Imputed	3755
-1	[-1] No answer	0
-2	[-2] Does not apply	0
-3	[-3] Implausible value	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
-7	[-7] Only available in less restricted edition	0
-8	[-8] Question not part of the survey program this year	0
-9	[-9] Don't want to answer [only CAWI]	0
-10	[-10] Break-off [only CAWI]	0

This variable designates imputations of item-nonresponse in the variable HGHINC (monthly household net income).