CZECH HOUSEHOLD PANEL SURVEY

Data Documentation

Waves 1 (2015) to 4 (2018)

Michaela Kudrnáčová (translation Jan Morávek)

Institute of Sociology, Czech Academy of Sciences

www.promenyceskespolecnosti.cz

promeny.spolecnosti@soc.cas.cz

English version 1.0 (10 October 2020), based on Czech version 5.0 (21 November 2019)

PROMĚNY ČESKÉ SPOLEČNOSTI

Table of Contents

List of Tables

List of Figures

Summary

Target population

Non-institutionalized population of the Czech Republic

Sampling unit

Household on common budget

Survey unit

In Wave 1, all members of the sampled households. In Waves 2 to 4, all members of households in which at least one panel member of the previous wave lived.

Sampling method

Two-stage stratified random sampling

Data collection agencies

MEDIAN and STEM/MARK

Data collection periods

Wave 1	Wave 2	Wave 3	Wave 4
7 Jul – 1 Dec 2015	20 Jun – 31 Oct 2016	21 Jun – 30 Oct 2017	20 Jun – 15 Oct 2018

Number of interviewed households

Wave 1	Wave 2	Wave 3	Wave 4
5,159	4,147	3,616	3,188

Household-level response and retention rates

Wave 1	Wave 2	Wave 3	Wave 4
42.1%	79.5%	84.7%	86.4%

Data collection modes

Standardized computer-assisted personal interviewing (CAPI), paper-and-pencil self-administered questionnaire (SAQ), computer-assisted web interviewing (CAWI)

Unit of analysis

Household and individual

Dataset content and structure

The main dataset, which is the subject of this documentation, contains variables from the following instruments: household CAPI questionnaire, individual and proxy CAPI questionnaires, adult self-completion, and self-completion for children aged 10–14 and 15–17 years, including identification, derived and interview context variables, and weights.

Time-use diary data and contact form data are available in separate files with their own documentation. All types of datasets have been created separately for each survey wave.

Data checks and cleaning

Checking value ranges and routing, cleaning basic sociodemographic variables, coding open-ended and semi-closed questions

Data weighting

Post-stratification weights for 16 sub-samples of respondents based on region of residence, sex, age and education. For time-use diary respondents, the basic weights were complemented by ones that ensure a uniform distribution of weekdays.

1. Czech Household Panel Survey

The Czech Household Panel Survey (CHPS) is a sample survey which repeatedly interviewed, in the years 2015–2018, a random sample of households living in the territory of the Czech Republic. 5,159 households were interviewed in the first wave of data collection (Wave 1) that took place from July to December 2015. In the following three years, households interviewed in the previous year were invited to take part. 3,188 households participated in the final Wave 4 in 2018.

For public relations purposes, the survey is referred to as *Proměny české společnosti* (Transformations of the Czech Society, www.promenyceskespolecnosti.cz).

The goal of the survey is to map the living conditions of Czech households in the long-term perspective, describe the dynamics of change in the life of households and individuals, and to relate the process of social change to relations and happenings within households.

CHPS is an inter-disciplinary study that relies on sociological, economic and political science approaches. It consists of five main thematic areas (*Table 1*).

Table 1: Thematic areas of the Czech Household Panel Survey (CHPS)

Family life, time use, health	Education and the labour market	Social stratification	Housing	Political participation and civil society
• Household relations • Work-family reconciliation • Child care • Income management • Gender attitudes • Time use • Health and mental wellbeing	• Educational pathways and opportunities • Educational aspirations • Employment, unemployment • Labour market mobility • Working conditions • Job satisfaction	• Social inequalities • Class structure • Social mobility • Income and wealth • Cultural capital	• Housing characteristics • Housing satisfaction • Preferred housing • Housing availability • Intergenera-tional transfer of housing ownership and attitudes	• Electoral participation and preferences • Party attachment • Political and civic attitudes • Civic engagement • Political discussions with family

The survey was co-organized by the Institute of Sociology, Czech Academy of Sciences, the CERGE-EI (a joint workplace of Charles University and the Economics Institute, Czech Academy of Sciences) and the Faculty of Social Studies, Masaryk University. Comprised of members of those institutions, the research team devised the survey's methodological plan and data collection instruments, selected fieldwork agencies through a public tendering procedure, collaborated with those agencies in implementing the survey's methodology, data checks and cleaning, processed the final datasets, and ensure that they were deposited in the Czech Social Science Data Archive.

The fieldwork was implemented by MEDIAN and STEM/MARK, two of the leading market research and opinion polling agencies in the Czech Republic. They are the founding members of SIMAR (Association of Market and Public Opinion Research Agencies) and members of ESOMAR (European Society for Opinion and Marketing Research). In addition to data collection, the agencies undertook the sampling process, collaborated in data checks and cleaning, and provided research methodology consultations.

The survey was funded by the Czech Science Foundation under grant GB14-36154G (Dynamics of Change in Czech Society).

2. Survey sample

2.1. Target population

The survey's target population comprises the non-institutionalized population of the Czech Republic irrespective of citizenship or ethnicity. A household was deemed eligible for the survey if it was able to be interviewed, i.e. at least some of its members spoke Czech or Slovak and interviewing was not prevented by their health condition.

The age eligibility limits vary across data collection instruments as they were designed for different age groups of household members (*Table 2*).

Table 2: Age limits for the populations of the different instruments

CAPI household questionnaire	no age limit
CAPI individual questionnaire	18+ years
CAPI proxy questionnaire	18+ years
Adult self-completion	18+ years
Older child self-completion	15–17 years
Younger child self-completion	10–14 years
Adult diary	18+ years
Child diary	10–17 years

2.2. Sampling unit

The sampling unit is household on common budget defined in line with the methodology of the Czech Statistical Office.

A household is comprised of individuals who usually reside in a given flat or family house and share a common budget, i.e. cover jointly the household's basic essential expenses including food, housing costs, other costs of running a household (including individuals such as children or incomeless persons who do not contribute to covering that expenditure themselves but their share is covered by other household members).

The survey collects information on all members of the sampled households. Household membership is defined by usual, rather than permanent residence.

Household members (if they share a common budget) include:

1 individuals who usually live in the dwelling (have spent most of their time outside work/school here during the past 6 months or, alternatively, are planning to reside here for at least 6 months);

2 lodgers and visitors who currently do not reside at another private address and who intend to reside in the surveyed dwelling for more than 6 months;

3 individuals who study or work outside their home for long periods, irrespective of the duration of their absence, if they do not have another private address and retain close ties with the surveyed household;

4 individuals who stay in hospitals or other institutions for long periods, if their absence does not exceed 6 months and they retain close ties with the surveyed household.

Students who reside outside their parental household over long periods are considered members of that household if they do not have any other private address at which they reside more often than in their parents' household, and they maintain financial ties with the parental household (in case of doubt, the decision rule was applied of whether the household gave the student at least CZK 12,000 over the past year).

2.3. Sampling the households

A two-stage stratified random sampling method was applied. Its design, however, was comparable with one-stage stratified cluster sampling. The design effects were further mitigated by the use of a large number of small primary sampling units.

The variables *strata* and *psu* in the dataset allow for correcting standard errors calculated under the assumption of simple random sampling.

The sample was constructed by the MEDIAN and STEM/MARK agencies. The sampling frame consisted of the current version of the Register of Census Districts and Buildings[1], which had been transformed into an address database (the number of times a building occurs in the database equals to the number of flats in that building). Sampling stages:

1 The territory of the Czech Republic was divided into 58 strata defined by NUTS 3 units (regions) and 5 municipality size categories (there were no municipalities in 12 of the theoretical 70 strata). The number of primary sampling units (addresses) selected in a stratum by means of systematic random sampling was proportional to the number of households residing in that stratum (according to the 2011 Population Census by the Czech Statistical Office) and expected response rate. The total number of primary sampling units was 1,275. Prior to the systematic random sampling procedure, the database units were sorted by attributes such as territorial unit number or building number.

[1] The Register of Census Districts and Buildings is a public list of buildings and their attributes maintained by the Czech Statistical Office. For more information, see https://www.czso.cz/csu/rso/-registr-scitacich-obvodu-a-budov

2 By means of simple random sampling, each primary sampling unit's address was assigned 16 additional addresses. Such addresses were no more than 500 meters from the primary address in municipalities with a population of at least 1000, whereas the distance from primary address was not limited in smaller municipalities. The resulting gross sample included 21,675 addresses (1,275 × 17).

The addresses sampled were subsequently issued to agencies' fieldwork executives (not interviewers themselves) to verify the existence and residential character of each address. Non-existing and non-residential addresses were excluded, and names indicated on letterboxes, doors etc. were recorded. This was done to reduce the number of non-existing and non-residential addresses during the main part of fieldwork and to allow for personalized advance mailing. For each primary sampling unit, 10 addresses were randomly selected from the verified gross sample. The resulting net sample included 12,750 addresses.

During fieldwork, 312 addresses that turned out to be vacant were substituted with unused addresses in the same primary sampling units randomly selected from the gross sample.

When more than one household on common budget was found in a dwelling, no procedure for selecting (some of) those households was prescribed. Instead, the interviewers were instructed to try and interview all such households.

2.4. Following rules

In Waves 2 to 4, the survey included such households in which at least one individual resided at the time of data collection who had been a member of a household interviewed in the immediately preceding wave ("original sample member").[2] Thus, when a household was not interviewed in Wave 2, its members were no longer invited in Waves 3 and 4; the same applies to households discontinued in Wave 3.

Only households residing outside institutions and in the territory of the Czech Republic were surveyed. Thus, when any or all members of a Wave 1 household moved to an institution or left the territory of the Czech Republic permanently, they ceased to be followed for interviewing purposes. However, they were again counted in as members of their former households in the following waves if they returned.

In Waves 2 to 4, the survey collected information on all individuals in the above-defined households, including joiners who had not belonged to the sampled households in the previous waves. In the following years, such joiners obtained the status of original household members. Thus, they were followed until the end of the survey, not only as long as they resided in the households of original Wave 1 members. CHPS did not implement the status of temporary sample members that exists in some other household panel surveys.

When an entire household moved to a new address, it was supposed to be interviewed in their new place of residence. It was not possible to replace the relocated household by a new household that was

[2] Two households were interviewed during Wave 2 in which no individual from the preceding wave currently resided, but where such an individual had been part of the household at a time in between both waves. In the meantime, the new members moved in with the original household members and the latter died.

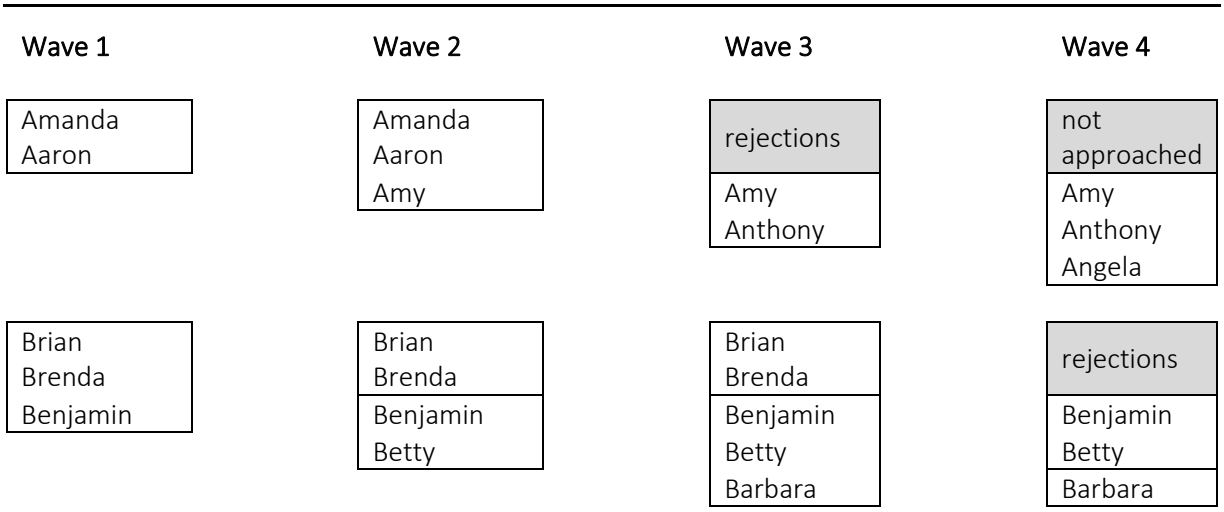
occupying the same dwelling. When several new households split off from a former household (e.g., when an adult child moved out their parents' place and started their own household), the goal was to interview both the household at the original address and the derived household at the new address. All newly formed households were to be surveyed even if none of them remained at the original address (i.e. all secondary households relocated to new addresses) or, in contrast, even if the split-off was confined to the original address (e.g., a multi-generational household split into two households residing in the same dwelling, each with their own budget).

The following two examples, as summarized in *Figure 1*, illustrate the principle of following rules applied to households and individuals.

Example A: Amanda and Aaron, who form a married couple, were interviewed in Wave 1. They lost their own housing before Wave 2 and moved in with their daughter Amy. Thus, the household of Amanda, Aaron and Amy was surveyed in Wave 2. Subsequently, Amanda and Aaron moved out of Amy's place, but they refused to participate in Wave 3. Amy's boyfriend Anthony moved in with her, and their household was interviewed. In Wave 4, Amanda and Aaron were not invited due to their non-participation in Wave 3. Daughter Angela was born into Amy and Anthony's household, and that household once again took part in the survey.

Example B: The household interviewed in Wave 1 consisted of a married couple, Brian and Brenda, and their son Benjamin. Before Wave 2, Benjamin moved in with his girlfriend Betty, and their new household took part in the survey. The remaining household of Brian and Brenda participated as well. In Wave 3, both Brian and Brenda's household and Benjamin and Betty's continued in the survey, the latter including Benjamin's sister Barbara who had moved in with them. Brian and Brenda no longer wanted to take part in Wave 4. Barbara moved out of Benjamin and Betty's household. Therefore, both Benjamin and Betty's household and the new single-member household of Barbara were approached, and both chose to participate in Wave 4.

Figure 1: Illustrating the rules for following individuals and households



3. Data collection

3.1. Interviewer training

Primarily those interviewers with long-term experience with one of the agencies, including surveying random samples and the CAPI method, were chosen to work on the CHPS. All CHPS interviewers underwent a dedicated training session at each wave.

One-day group sessions for Wave 1 took place in selected cities of the Czech Republic during June 2015. Several interviewers were briefed on an individual basis. For Waves 2 to 4, approximately 80% of the interviewers attended a one-day group session in May and June of each year, before fieldwork commenced; the remaining interviewers were again briefed individually (either in person or through an e-learning course).

In all waves, the interviewer training focused on the following:

- relevance and mission of CHPS; the academic institutions behind it,
- time schedule of CHPS in 2015–2018,
- household sampling procedure,
- definitions of household and household membership,
- data collection instruments and how to work with them,
- time-efficient ways of in-home interviewing (handing out the questionnaires for self-completion while another household member is being interviewed face-to-face),
- questionnaire sections that are more demanding or crucial for the next steps in the interview (e.g., examining household relations and checking them in subsequent waves, or cognitive ability testing in Wave 2),
- communication with respondents (explaining the sampling method, ensuring confidentiality, the importance of taking part, irreplaceability of household in the sample),
- strategies to increase response and conversion rates,
- respondent and interviewer incentives,
- fieldwork quality control mechanisms,
- fieldwork monitoring.

In Waves 2 to 4, the briefings also covered specific issues of repeated interviewing, and in particular:

- finding and contacting relocated households and individuals,
- explaining the importance of repeated participation in the survey,
- accounting for joiners and leavers,
- dependent interviewing in the household and CAPI individual questionnaires.

The interviewer training in Wave 1 was preceded by a “training of trainers” organized by project team members from the Institute of Sociology. In this workshop, agency staff and the research team discussed and refined the key points of the upcoming interviewer briefing.

3.2. Piloting

In each wave, the main part of the fieldwork was preceded by a pilot study with the aim of testing and fine-tuning the survey's parameters, and in particular:

- the data collection instruments (comprehensibility, time requirements, the functioning of CAPI scripts),
- respondent and interviewer incentives,
- fieldwork quality control and monitoring mechanisms.

From 24 April to 19 May 2015, Wave 1 pilot survey took place. 302 households were interviewed following a sampling procedure analogous to that implemented for the main part of fieldwork.

The Wave 2 pilot took place from 13 April to 9 May 2016. In addition to the 234 households that had taken part in the 2015 pilot, 71 new households were interviewed (a total of 305 households). The new households were selected within a 500-meter distance from the original Wave 1 households that had refused to continue or could not be reached.

The pilot exercise for Wave 3 was conducted from 27 March to 24 April 2017 with 272 of the households that had been interviewed in Wave 2. Wave 4 piloting took place from 26 March to 24 April 2018 with 258 households from Wave 3 sample.

3.3. Advance mailing

In Wave 1, the following documents were mailed one week in advance of each data collection stage:

- an introductory letter to the respondents requesting their cooperation, and a respondent leaflet with contact information (phone, e-mail, website) – by mail,
- a letter to local mayors requesting them to inform the population – by mail and by e-mail,
- a letter to regional police directorates with information about interviewers' activities – by mail and by e-mail.

In Waves 2 to 4, a request for repeated participation was mailed to the households in advance, including information about lottery winners and a leaflet with selected results of the preceding wave and contact information.

3.4. Fieldwork

The fieldwork for **Wave 1** data collection took place from 7 July to 1 December 2015. It was divided into four monthly stages, with batches of addresses issued to interviewers at each stage. More than half of the addresses (about 60%) were distributed during the first two months so that the interviewers could repeat contact attempts during a longer time period and increase the chances of contacting even hard-to-reach households. Due to higher rates of non-contact during summer holidays (July, August), approximately two-thirds of the households were interviewed from September to December 2015.

Of the 13,269 addresses issued, 5,159 households were interviewed and successfully validated by quality checks (42.1% of the eligible addresses, see *Table 3* for details).

301 interviewers participated in the data collection effort with at least one household interviewed. Each interviewer was assigned 1 to 6 primary sampling units per data collection stage. In the last two stages, the primary sampling units of low-activity interviewers were reassigned to other interviewers. On average, 15 households were surveyed by every interviewer.

The sample of 12,750 addresses was split equally between MEDIAN and STEM/MARK. Both agencies exhibited similar response rates (2,562 households were interviewed by MEDIAN and 2,597 by STEM/MARK).

Table 3: Field outcomes for the household sample, Wave 1

Addresses issued	13,269	*)
Eligible households	12,265	
Complete interview, validated by quality control	5,159	42.1%
Refusal	4,216	
Interview incomplete or arranged for after fieldwork deadline	149	
Non-contact (of entire household or adult member)	2,265	
Household not visited	252	
Interviewed, excluded by quality control	224	
Ineligible households	1004	7.3%
Unable to take part – no member speaks Czech	14	
Unable to take part – health problems, old age	286	
Address not found	299	
Vacant/non-residential	405	

*) During fieldwork, 207 households on common budget living at the same address were added to the original sample of 12,750 addresses (when more than one household on common budget was found in a dwelling), and 312 addresses were randomly sampled to replace those that turned out during fieldwork to be vacant or non-residential.

The fieldwork **in Wave 2** took place from 20 June to 31 October 2016. Household addresses were issued to interviewers in three stages of approximately six weeks each. About 80% of the addresses were distributed in the first two stages. The households were assigned in a sequential manner, depending on the stage in which they had been interviewed in Wave 1. Interviewing a household after less than 10

months of Wave 1 was not allowed. More than half of the households (55%) were interviewed from June to August, one-third in September, and the remaining 12% in October.

The households interviewed and validated in Wave 1 (5,093 addresses) were invited to participate in Wave 2. During fieldwork, the list of addresses issued was expanded to include 171 split-off households originating from Wave 1 households. In other words, for 171 households interviewed in Wave 1, it was found that some members had left and formed new households. Only in 16 cases of such split households was an interview in the primary household successfully followed by an interview in the secondary household as well.

A total of 4,147 households were interviewed and validated in Wave 2 (79.5% of eligible addresses, see *Table 4* for details). Of the 4,147 households participating, 137 (3.3%) were interviewed at a different address than in Wave 1 (change of address includes relocation to another dwelling at the same address). Relocation applied either to the entire household or to selected members only. 47% of relocated households were successfully interviewed.

208 interviewers participated in data collection with at least one household interviewed. On average, 18 households were interviewed by every interviewer. 3,727 households took part in face-to-face interviewing during both waves, and 84% of them were interviewed by the same interviewer. The MEDIAN and STEM/MARK agencies contributed to the fieldwork effort comparably. (2,050 households were interviewed by MEDIAN and 2,097 by STEM/MARK).

Table 4: Field outcomes for the household sample, Wave 2

Addresses issued	5,264		
Interviewed in W1	5,093	*)	
New split-off households	171		
Eligible households	5,217		
Complete interview, validated by quality control	4,147		79.5%
Refusal	310		
Interview incomplete or arranged for after fieldwork deadline	7		
Non-contact (of entire household or adult member)	101		
Household did not respond to CAWI invite (W1 CAWI households only)	227		
Household not visited	256		
Change of address, not interviewed	163		
Interviewed, excluded by quality control	6		
Ineligible households	47		0.9%
Unable to take part – no member speaks Czech	0		
Unable to take part – health problems, old age	45		
Address not found	2		

*) In Wave 2, 67 of the 5,159 households interviewed using CAWI in 2016 were not approached because no functional e-mail address was available for them. In contrast, one household that had been excluded from Wave 1 data due to interviewer errors did take part in Wave 2.

Wave 3 data collection took place from 21 June to 30 October 2017. To maintain at least 10 months' interval between instances of interviewing, the addresses were issued to interviewers in batches depending on their last year's interviewing date. STEM/MARK divided the fieldwork into two stages and MEDIAN into three. Almost two-thirds of the households (64%) were interviewed from June to August, 29% in September, and the remaining 7% in October. Administration of the CAWI mutations of self-completion questionnaires and diaries for adults and children continued throughout November in those households that had not been interviewed within standard deadline due to errors in CAWI settings (some CAWI sessions had been terminated before redirection to the above-mentioned instruments). These households were requested to complete a follow-up interview during November; approximately 20% of CAWI households completed their interview in this way.

Only the addresses interviewed and successfully validated in Wave 2 were issued for interviewing. The interviewers received a total of 4,303 addresses to interview, of which 156 were split-off households. Both the primary and the secondary household were successfully interviewed in 11 cases.

3,616 households were interviewed and validated in Wave 3 (84.7% of eligible addresses, see *Table 5* for details). For 130 of them (3.6%), interview took place at a different address than in Wave 2. The success rate of interviewing at new addresses was 46%.

170 interviewers contributed at least one household, with an average of 19 households per interviewer. Interviewer continuity was maintained for 76% of the 3,195 households interviewed face-to-face in all three waves. The households interviewed were again split evenly between MEDIAN (1,794) and STEM/MARK (1,822).

Table 5: Field outcomes for the household sample, Wave 3

Addresses issued	4,303	
Interviewed in W2	4,147	
New split-off households	156	
Eligible households	4,269	
Complete interview, validated by quality control	3,616	84.7%
Refusal	128	
Interview incomplete or arranged for after fieldwork deadline	29	
Non-contact (of entire household or adult member)	79	
Household did not respond to CAWI invite	108	
Household not visited	149	
Change of address, not interviewed	151	
Interviewed, excluded by quality control	9	
Ineligible households	34	0.8%
Unable to take part – no member speaks Czech	0	
Unable to take part – health problems, old age	32	
Address not found	2	

Wave 4 fieldwork was conducted between 20 June and 15 October 2018. As in previous years, addresses for face-to-face interview and on-line data collection were issued in batches so that a household would participate no earlier than 10 months after Wave 3 interview. STEM/MARK divided the fieldwork into two stages and MEDIAN into three. By the end of August, data was obtained from 65% households, and the remaining 35% were interviewed in September and October.

The sample issued in Wave 4 included all households that participated in the previous year (3,616). In the course of the fieldwork, 118 split-off households were added to the sample. Interviews with both primary and secondary household were secured in 10 cases.

In Wave 4, data was collected and validated from 3,188 households resulting in 86.4% household response rate. Out of these, 102 households moved to a different address since their participation in Wave 3. The success rate of interviewing household at a new address is estimated at 48%.

157 interviewers contributed to Wave 4 data collection with at least one interviewed household (an average of 18 households per interviewer). Face-to-face interview was carried out in 2,829 households in all four waves. 71% of them were interviewed by the same interviewer, at least one change of interviewer occurred in the remaining cases.

MEDIAN and STEM/MARK again had a comparable share in fieldwork (1,556 and 1,632 households).

Table 6: Field outcomes for the household sample, Wave 4

Addresses issued	3,734	
Interviewed in W3	3,616	
New split-off households	118	
Eligible households	3,688	
Complete interview, validated by quality control	3,188	86.4%
Refusal	128	
Interview incomplete or arranged for after fieldwork deadline	5	
Non-contact (of entire household or adult member)	29	
Household did not respond to CAWI invite	108	
Household not visited	109	
Change of address, not interviewed	111	
Interviewed, excluded by quality control	10	
Ineligible households	46	1.2%
Unable to take part – no member speaks Czech	0	
Unable to take part – health problems, old age	44	
Address not found	2	

3.5. Approaching the households

In **Wave 1,** the interviewers were instructed to undertake at least six attempts to approach a household (distributed across different times of the day and week). Four visits were to be undertaken in the month following assignment, and two were to be deferred to maximize the chances of contacting the household. The first attempt to approach the household was due in the week following the assignment of the primary sampling unit. Interviewer incentives were designed in a way to motivate closing a primary sampling unit during the month following assignment.

The first contact with the household took the form of a personal visit by the interviewer, whereas phone communication was only allowed after establishing the first contact. An interviewer who failed to reach the household could leave in the letterbox a business card with the date and time of the visit and their contact information so that the household could itself get in touch with the interviewer.

The introductory letter and respondent leaflet mailed to sampled households contained phone numbers and e-mail addresses for direct contact with the fieldwork agencies or the research team. A dedicated respondent section was also created at the survey website, www.promenyceskespolecnosti.cz.

Interviewers received copies of the introductory letter and respondent leaflet for presentation when approaching the households. All interviewers attended training that covered certain communication principles with a view to maximizing the response rate (relevance of CHPS, sampling procedure, irreplaceability of household in the survey, ensuring confidentiality, financial incentives, etc.)

For fieldwork monitoring purposes, the interviewers recorded the days and outcomes of each contact attempt. When a household they approached refused to take part, they recorded its refusal reasons. A distinction was made between "hard" refusals, which resulted in terminating contact with the household, and "soft" refusals (e.g., bad timing), where the interviewer attempted to approach the household again. *Table 7* shows the final results for refusal reasons in each wave.

Since September 2015, several additional measures were taken to increase the chances of contacting hard-to-reach households and converting refusals. Public records were searched for phone numbers of unreached households. Subsequently, phone operators contacted those households and arranged the interviewer's visit if the household agreed to take part in the survey. In October and November, refusing or repeatedly non-contacted households were invited to participate online (in the CAWI instrument) by means of a letter with a unique access code to the survey portal.

In **Waves 2 to 4,** interviewers were instructed to make at least four attempts at contacting a household, including two attempts in the month following assignment and two attempts at a later time period. To establish contact, they were advised to use the phone or e-mail information obtained from household members during previous waves. In Wave 2, the interviewer incentives were designed to motivate approaching households in the first week after assignment and closing the assignment within the deadline of the current fieldwork stage. In Waves 3 and 4, such a measure was no longer necessary thanks to high response rates. Again, the interviewers recorded every contact attempt, its circumstances and outcomes for monitoring purposes.

The advance letters sent in Waves 2 to 4 contained the interviewer's phone number so that the household could itself arrange the interview. Like in Wave 1, the interviewers were provided with business cards and copies of the introductory letter and respondent leaflet.

In case an entire household or some of its members relocated, the interviewer was instructed to inquire from other household members or neighbours, or using Wave 1 phone and e-mail contact information, about the new address or contact information of the relocated individuals, and to report relocation to their agency. When the new address was outside the interviewer's reach, it was assigned to a new interviewer. A special bonus was due for obtaining a relocated household's contact information, no matter if the interviewer interviewed that household themselves. An additional bonus was awarded for surveying the relocated household successfully. The greatest obstacle to interviewing relocated members of split households was posed by the fact that other members of the original household did not know or were unwilling to provide the former member's place of residence or contact information.

Except for offering a CAWI conversion, no special measures to increase response rates were taken in Waves 2 to 4.

Table 7: Reason for refusal

	Wave 1	Wave 2	Wave 3	Wave 4
Not enough time (household members constantly busy)	10.4%	13.5%	28.9%	35.9%
Bad timing, extraordinary situation in household (just leaving, illness, visit etc.)	2.3%	5.5%	7.8%	6.3%
No interest/trust in surveys in general	52.9%	39.7%	6.3%	5.5%
Recently responded to a survey	0.3%	0.0%	0.0%	0.0%
Bad prior experience with surveys	1.6%	1.3%	0.8%	0.0%
No interest in this survey's topics	5.3%	7.1%	13.3%	4.7%
No trust in this survey (not enough information, do not like the topic etc.)	3.1%	1.0%	1.6%	0.0%
Concerns about confidentiality/misuse of information/sharing personal data	5.9%	2.6%	4.7%	3.9%
Insufficient incentives – low reward etc.	0.5%	1.0%	0.0%	3.1%
Afraid of interviewer (afraid to open door, afraid of door-to-door selling)	1.1%	0.3%	0.0%	1.6%
A household member "forbade" taking part	4.4%	8.7%	6.3%	3.1%
Other (e.g., household itself notified agency of its refusal to take part)	1.8%	0.6%	5.5%	0.0%
No reason given	6.8%	11.9%	1.6%	0.0 %
Not available	3.6%	6.8%	23.4%	35.9 %

3.6. CAWI interviewing

In Wave 1, a CAWI mutation of the interview was prepared as one of the strategies to convert refusals or repeatedly non-contacted households. The online questionnaires were completed by 589 households in Wave 1 (11% of the total number of households interviewed). The CAWI conversion rate was 8%.

From Wave 2, the households previously interviewed online were no longer invited by interviewers to participate in face-to-face interviewing and, instead, they were immediately sent a CAWI invite via e-mail or a special letter. The CAWI retention rate (i.e. the share of previous wave's CAWI households that took part in CAWI in the current wave) reached 49% in Wave 2 and 70% in Wave 3 and Wave 4. Only exceptionally, upon their own request, were such households interviewed face-to-face.

From Wave 2, only the MEDIAN agency employed the CAWI method as a strategy to convert households refusing face-to-face interviewing. STEM/MARK did not resort to conversion thanks to high face-to-face retention rates. The CAWI conversion rates (i.e. shares of households that accepted the invitation to convert from CAPI to CAWI) ranged between 25% and 30%.

Table 8: Numbers of households surveyed and retention/conversion rates by data collection mode

	Wave 1		Wave 2		Wave 3		Wave 4	
CAPI retention	-		3,743		3,222		2,866	
	-		80.7%		84.9%		86.8%	
CAPI conversion	-		1		2		0	
	-		-		-		-	
CAPI total	**4,570**	88.6%	**3,744**	90.3%	**3,224**	89.2%	**2,866**	89.9%
CAWI retention	-		294		292		284	
	-		48.8%		69.5%		70.3%	
CAWI conversion	589		108		100		38	
	8.3%		25.7%		24.9%		30.9%	
CAWI total	**589**	11.4%	**402**	9.7%	**392**	10.8%	**322**	10.1%

3.7. Fieldwork monitoring

Interviewers used the contact forms to record the days, outcomes and other circumstances of their contact attempts made at each address. Using the CAPI portal, they transmitted this information, along with a list of interviews completed, as soon after approaching the household as possible. Based on this information, the agencies prepared weekly fieldwork status reports for the research team.

Fieldwork supervisors engaged in regular communication with every interviewer about their work thus far and data collection targets. In Wave 1, due to a lower response rate in initial stages, the frequency of telephone communication between supervisors and interviewers was increased to at least two calls per week during the last two months.

Complete contact form data is available in separate datasets.

3.8. In-home interviewing

The recommended in-home interviewing strategy was to arrange the appointment for a time when as many members were present in the household and to complete as many instruments as possible during one session. However, interviewers were able to revisit households to finish interviews, interview additional individuals or pick up self-completions.

The average time spent in a household was 110 minutes in Wave 1 and 120 minutes in Waves 2 to 4. In the opening part of the session, the interviewer presented the survey, the household sampling procedure, what is requested of respondents, data protection arrangements and related respondent rights, the respondent incentive system, etc. In addition, the respondents were instructed on self-administered instruments (adult/child questionnaires and diaries) to make the interviewing session as efficient as possible. The optimal procedure for multiple-member households was for other members to complete their self-administered questionnaires while the interviewer engaged in face-to-face interviewing.

Each self-completion instrument was supplied with a ballpoint pen. To facilitate the interviewing session, the interviewers were carrying colouring books and crayons for entertaining younger children while adults were being interviewed.

The interviewer concluded each session by collecting contact information (e-mail addresses and phone numbers) from every household member using a dedicated form, including informed consent, and preparing the necessary paperwork for the payment of household incentives. In addition, the interviewer checked the formal accuracy of self-completed instruments and asked the respondents for any corrections or additions. A thank-you card was handed out to the participating household.

3.9. Respondent and interviewer incentives

For its participation in the survey, each household obtained a financial incentive of CZK 500–1000, depending on household size. In Wave 1, a bonus of CZK 100 was awarded to households that completed all instruments possible. Due to the complicated process of administering the bonus and checking accurate disbursement, the bonus was discontinued in the subsequent waves. Every household was able to forward its financial incentive to a charity. Before the start of fieldwork in Waves 2 to 4, sixteen of the preceding wave's participating households were randomly drawn and obtained shopping vouchers. In Waves 2 and 3, completion of child questionnaires was rewarded with small gifts (mini board games, reflective stickers) to maximize the response rate for this type of instruments.

The interviewer incentive system was based on the number of instruments completed per household. The basic incentive was awarded for completing the household CAPI questionnaire and the first CAPI individual questionnaire. A unit rate defined separately for each instrument was paid for every additional individual CAPI, proxy CAPI or self-administered questionnaire. Incentives were paid only for households in which at least half of the members were interviewed, and at least half of the possible instruments were completed.

Throughout the different waves, the interviewer incentive system was adjusted to reflect specific fieldwork needs. In Waves 1 and 2, bonuses were awarded for early completion of a primary sampling

unit assignment. In Wave 1, the rate per household interviewed was increased from the beginning of October. After the fieldwork was completed in Waves 2 to 4, special bonuses were awarded to the best-performing interviewers in accordance with individual evaluations of their work. In Waves 3 and 4, the child self-completion questionnaire and diary rates were increased. From Wave 2, the interviewers obtained bonuses for delivering the contact information of relocated or split households.

3.10. Fieldwork quality control

To verify interviews, two basic mechanisms were employed in all waves of the survey:

- listening to the audio recordings of selected parts of the interview made with the respondent's consent,
- follow-up checks by telephone, e-mail or mail to verify basic information about the household and the interviewing procedure.

These checks served to validate households interviewed face-to-face, with at least one check implemented for each household. Listening to audio recordings was done for 85–95% of households, depending on wave.[3] Follow-up checks by telephone, e-mail or mail were applied to 94% of households in Wave 2. From Wave 2, follow-up checks were only done by telephone, covering 53% of households in Wave 2, 79% in Wave 3 and 72% in Wave 4. When a check gave rise to doubts about the interviewer's conduct in a particular case, the other type of check was performed. Second checks were also applied to all other households of interviewers for whom any error was detected. When the interviewer was suspected of having conducted the interview with a household other than prescribed, a fieldwork supervisor checked the situation in person in the field.

In Wave 1, the above two mechanisms for verifying interviews were complemented by checking visits to the addresses (whether or not they were ultimately interviewed) by means of unique codes printed on the introductory letters. This check was merely supplemental since not all households were able or willing to provide the interviewer with the letter they had received.

The above checks were not applied to CAWI interviews because those were based on unique access codes delivered by mail directly to the households sampled, without an interviewer's intervention. Verification of the accuracy of CAWI completion was included in the data checks process (focusing, for example, on cross-wave consistency of sociodemographics).

3.11. Panel maintenance

Both during and beyond the data collection process, the respondents were able to retrieve information about the course of the survey at the dedicated website, www.promenyceskespolecnosti.cz. Published were fieldwork status updates as well as short analytical outcomes, some of which were further publicized through mass media. The survey's Facebook profile served to communicate information and share links to the website. During December, all households participating in the most recent wave obtained an electronic Christmas note with information about successful conclusion of the fieldwork.

[3] Covered were always 100% of the recordings where the respondent gave consent and recording was not prevented by a technical problem.

4. Data collection instruments

4.1. Types of instruments

Using standardized questionnaires, the data was collected by means of computer-assisted personal interviewing (CAPI questionnaires) or paper-and-pencil self-completion (self-administered questionnaires and time-use diaries). For the purposes of online data collection, all instruments were converted to a CAWI form as well.

The **household CAPI** was conducted with one household representative, typically the person most familiar with the household's housing and financial situation. An additional member could participate in the interview if he/she was better informed about an issue or able to correct/complement some answers or complete them him/herself. The questions related both to the household as a whole and to its individual members, irrespective of age.

The **CAPI individual questionnaire** targeted all adult members of the household. When a member was unable to complete the questionnaire (absence, health condition, etc.), another adult member could answer basic questions for him/her using the **CAPI proxy questionnaire.** The share of proxy questionnaires in the total number of adult CAPI questionnaires oscillated around 5%, depending on wave. *Table 9* indicates the average duration of each type of CAPI interview.

Showcards were prepared to accompany selected questions of the household, individual and proxy CAPI. The use of showcards is indicated under the applicable questions in the questionnaire.

Table 9: Average duration of a CAPI interview, in minutes

	Wave 1	Wave 2	Wave 3	Wave 4
Household CAPI	16.5	9	11.5	11
Individual CAPI	28	40.5	28.5	31
Proxy CAPI	8	9	9.5	7.5

The **adult self-completion questionnaires** were distributed to all household members aged 18 years or older, the **child self-completion questionnaires** to children aged 10–17. Although both instruments had some questions in common, their contents were largely different. Two versions of child self-completion were prepared for Waves 1 and 2, namely for the age groups of **10–14** and **15–17 years**. The number of questions included in and formatting of the young children's version were less demanding of the respondents. Due to a low number of child respondents, a unified version analogical to the adult version was implemented from Wave 3. The unified child questionnaire, however, had a basic section for all respondents and an extension for older children only.

The **diary** covering time use on the previous day was also self-administered. The diary came in two versions with different time use categories, one **for children** aged 10–17 years and another one **for adults**.

The self-completion questionnaires and diaries were distributed to the respondents for independent completion, and the interviewer only explained how to work with them. Children were instructed to

complete their questionnaires without parental supervision whereas, in case of difficulty, parents were allowed to assist their (especially younger) children with the diary.

4.2. Instrument administration

No set order of instrument completion was prescribed for face-to-face interviewing, except the rule that the household CAPI should precede the CAPI individual questionnaires. When the household CAPI was completed, the interviewing script presented a list of all instruments open for completion in the household. The system notified the interviewer on how many questionnaires could be completed in the household and how many were left until the household could be deemed successfully surveyed.

Respondents were able to complete their self-administered questionnaires and diaries both before and after their individual CAPI. Typically, the respondents worked on the self-completions while the interviewer was conducting a CAPI with another household member. In some cases, the interviewer distributed the self-completions and related instructions during the first contact with the household while arranging the CAPI for a later date. There were also situations when the self-administered instruments were completed after the interviewer left the household, for later pick-up.

In the CAWI mutation, the household questionnaire constituted the first section, followed by the individual questionnaire that could be completed one per household. Then, the respondents were taken to a portal with links to self-administered questionnaires and diaries for the different household members. Those were accessible both from the portal and through links e-mailed to the respective household members. Respondents were able to return to the portal later.

In the absence of an instrument completion hierarchy, each respondent exhibits a unique combination of instruments designed for his/her age group (except the combination of individual CAPI and proxy CAPI). Only the CAWI households were subject to the hierarchy described above.

4.3. Response rates by instrument

Households in which at least 50% of members were interviewed and at least 50% of instruments were completed, including the household CAPI and at least one CAPI individual questionnaire, were considered successfully surveyed. The data also includes partially surveyed households in which at least the CAPI household questionnaire was completed. This applies to 174 households in Wave 1, 85 in Wave 2, 112 in Wave 3 and 68 in Wave 4.

Table 10: Numbers of instruments completed, Wave 1

Household CAPI, household-level variables	5,159
Household CAPI, individual-level variables	13,083
of it:	
Adults	10,476
Children 10 to 17 years	1,079
Children under 10 years	1,528
Individual and proxy CAPI	7,605
of it:	
Individual	7,118
Proxy	487
Adult self-completion	8,131
Child self-completion	866
of it:	
15–17 years	306
10–14 years	560
Adult time-use diary	7,955
Child time-use diary	804

Table 11: Response rates by instrument, total and by household type (*hhtype*), Wave 1

	Total	One-person household	Married / same-sex partnership / cohabiting couple	Single parent with adult children	Single parent with children under 18 years	Parents with adult children	Parents with children under 18 years	Other
Individual CAPI excluding proxy	67.9%	99.6%	71.8%	63.1%	85.4%	52.1%	67.9%	50.0%
Individual CAPI including proxy	72.6%	99.6%	75.6%	69.5%	88.7%	58.4%	74.0%	54.9%
Adult self-completion	77.6%	78.0%	77.6%	75.3%	83.2%	75.0%	81.1%	73.3%
Child self-completion	80.3%	-	-	-	79.9%	-	79.2%	89.1%
Adult time-use diary	75.9%	77.1%	76.1%	72.2%	81.0%	74.5%	79.0%	70.6%
Child time-use diary	74.5%	-	-	-	74.7%	-	73.6%	81.8%

Table 12: Numbers of instruments completed, Wave 2

Household CAPI, household-level variables	4,147
Household CAPI, individual-level variables	10,498 *)
of it:	
Adults	8,394
Children 10 to 17 years	886
Children under 10 years	1,218
Individual and proxy CAPI	5,603
of it:	
Individual	5,270
Proxy	333
Adult self-completion	6,561
Child self-completion	602
of it:	
15–17 years	209
10–14 years	393
Adult time-use diary	6,723
Child time-use diary	602

*) Leavers (*rstat* = 5) are excluded from household size data.

Table 13: Response rates by instrument, total and by household type (*hhtype*), Wave 2

	Total	One-person household	Married / same-sex partnership / cohabiting couple	Single parent with adult children	Single parent with children under 18 years	Parents with adult children	Parents with children under 18 years	Other
Individual CAPI excluding proxy	62.8%	99.8%	66.6%	55.5%	84.4%	47.4%	61.4%	43.5%
Individual CAPI including proxy	66.8%	99.8%	69.8%	62.9%	88.8%	53.6%	65.5%	47.5%
Adult self-completion	78.2%	83.9%	78.3%	72.4%	83.4%	74.5%	80.5%	74.6%
Child self-completion	67.9%	-	-	-	61.5%	-	70.7%	58.8%
Adult time-use diary	80.1%	82.8%	81.1%	75.3%	84.4%	77.7%	81.9%	76.5%
Child time-use diary	67.9%	-	-	-	62.9%	-	70.0%	61.9%

Table 14: Numbers of instruments completed, Wave 3

Household CAPI, household-level variables	3,616
Household CAPI, individual-level variables	9,183 *)
of it:	
Adults	7,357
Children 10 to 17 years	787
Children under 10 years	1,039
Individual and proxy CAPI	4,870
of it:	
Individual	4,635
Proxy	235
Adult self-completion	5,839
Child self-completion	644
of it:	
15–17 years	233
10–14 years	411
Adult time-use diary	5,966
Child time-use diary	625

*) Leavers (*rstat* = 5) are excluded from household size data.

Table 15: Response rates by instrument, total and by household type (*hhtype*), Wave 3

	Total	One-person household	Married / same-sex partnership / cohabiting couple	Single parent with adult children	Single parent with children under 18 years	Parents with adult children	Parents with children under 18 years	Other
Individual CAPI excluding proxy	63.0%	99.9%	66.3%	54.3%	86.0%	47.9%	62.2%	45.2%
Individual CAPI including proxy	66.2%	99.9%	69.2%	59.1%	88.7%	52.6%	65.7%	48.5%
Adult self-completion	79.4%	88.4%	77.4%	75.8%	86.7%	78.0%	80.3%	76.0%
Child self-completion	81.8%	-	-	-	74.5%	-	83.1%	82.0%
Adult time-use diary	81.1%	88.3%	80.5%	79.2%	86.0%	79.6%	80.8%	78.6%
Child time-use diary	79.4%	-	-	-	68.9%	-	80.6%	84.3%

Table 16: Numbers of instruments completed, Wave 4

Household CAPI, household-level variables	3,188
Household CAPI, individual-level variables	8,093 *)
of it:	
Adults	6,532
Children 10 to 17 years	699
Children under 10 years	862
Individual and proxy CAPI	4,222
of it:	
Individual	4,021
Proxy	201
Adult self-completion	5,132
Child self-completion	572
of it:	
15–17 years	212
10–14 years	360
Adult time-use diary	5,391
Child time-use diary	562

*) Leavers (*rstat* = 5) are excluded from household size data.

Table 17: Response rates by instrument, total and by household type (*hhtype*), Wave 4

	Total	One-person household	Married / same-sex partnership / cohabiting couple	Single parent with adult children	Single parent with children under 18 years	Parents with adult children	Parents with children under 18 years	Other
Individual CAPI excluding proxy	61.6%	99.9%	65.3%	54.0%	78.3%	46.8%	60.5%	44.2%
Individual CAPI including proxy	64.6%	99.9%	67.6%	59.9%	80.6%	51.8%	63.3%	48.1%
Adult self-completion	78.6%	85.5%	76.3%	75.7%	86.0%	78.1%	79.0%	77.5%
Child self-completion	81.8%	-	-	-	75.3%	-	82.2%	85.9%
Adult time-use diary	82.5%	87.7%	81.1%	80.8%	85.3%	81.7%	82.1%	83.6%
Child time-use diary	80.4%	-	-	-	74.1%	-	81.2%	81.2%

The datasets include variables indicating the numbers and shares of instruments completed in a household (see Section 12, *Interview context variables*).

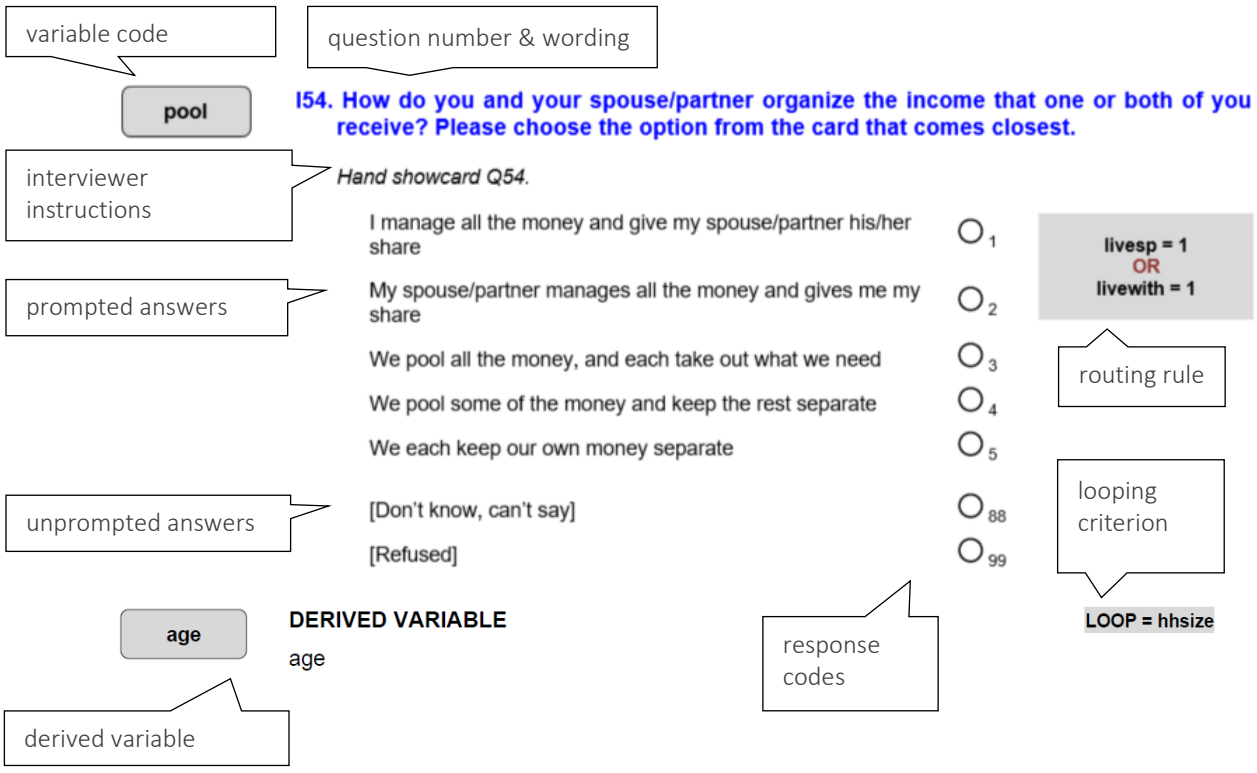
4.4. Instrument documentation

Variable labels in the dataset are a short representation of the question's wording. Full wording including answer options can be found in the different instruments. Value labels typically include the full wording of the option. Only a few extremely long wordings were shortened for the value labels.

CAPI data collection relied on software solutions of MEDIAN and STEM/MARK, which implemented the instruments in ways comparable both functionally and visually. The layout and formatting of the CAPI questionnaires published differs from those in the software environment.

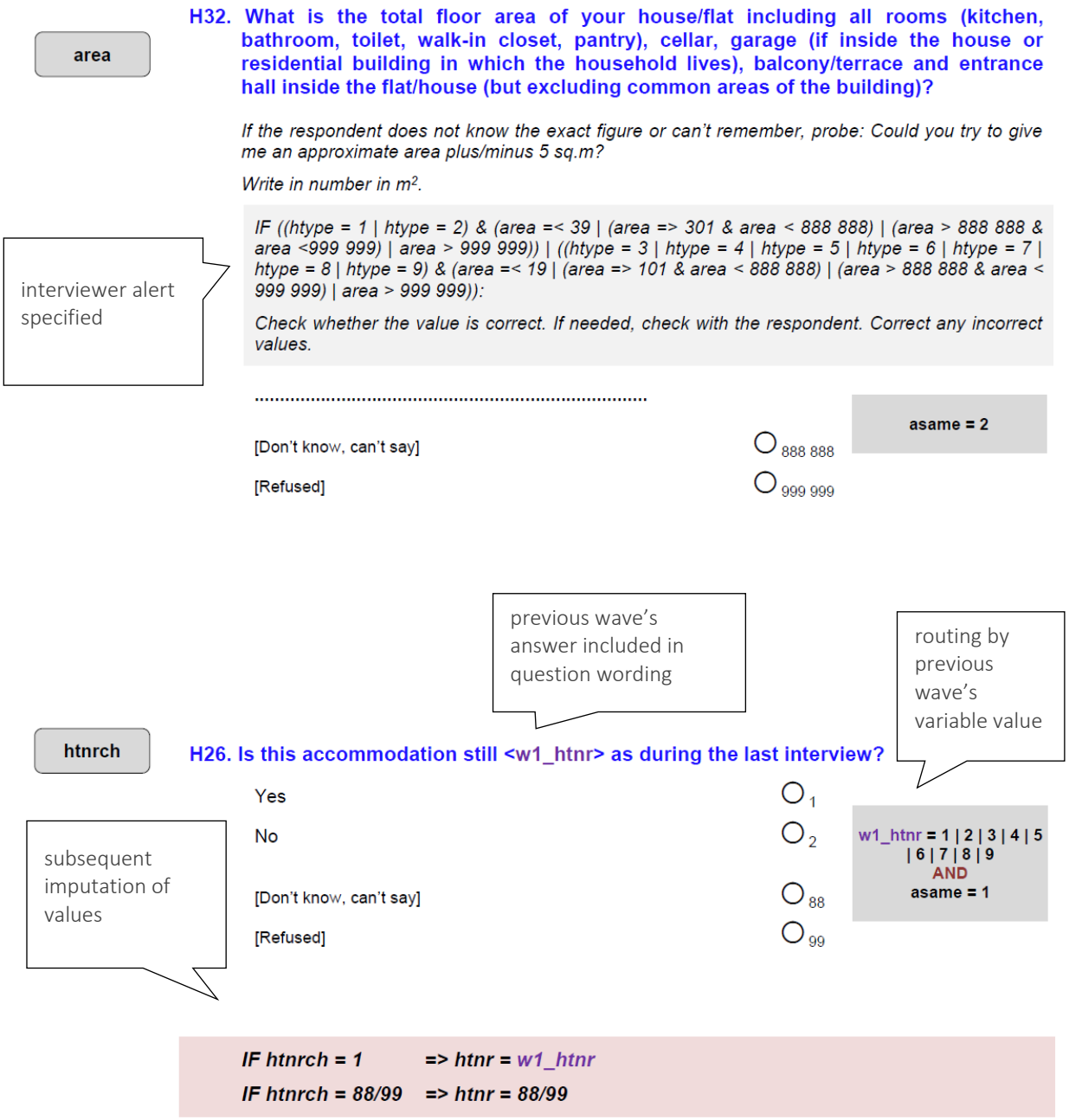
The CAPI questionnaires published contain variable codes, numbers and wordings, interviewer instructions, routing rules, prompted ("valid") answers and unprompted ("invalid") answers. The unprompted answers are enclosed by square brackets […]. The routing rule area also contains a question looping criterion when applicable.

Figure 2: Basic constituent parts of the CAPI questionnaires published



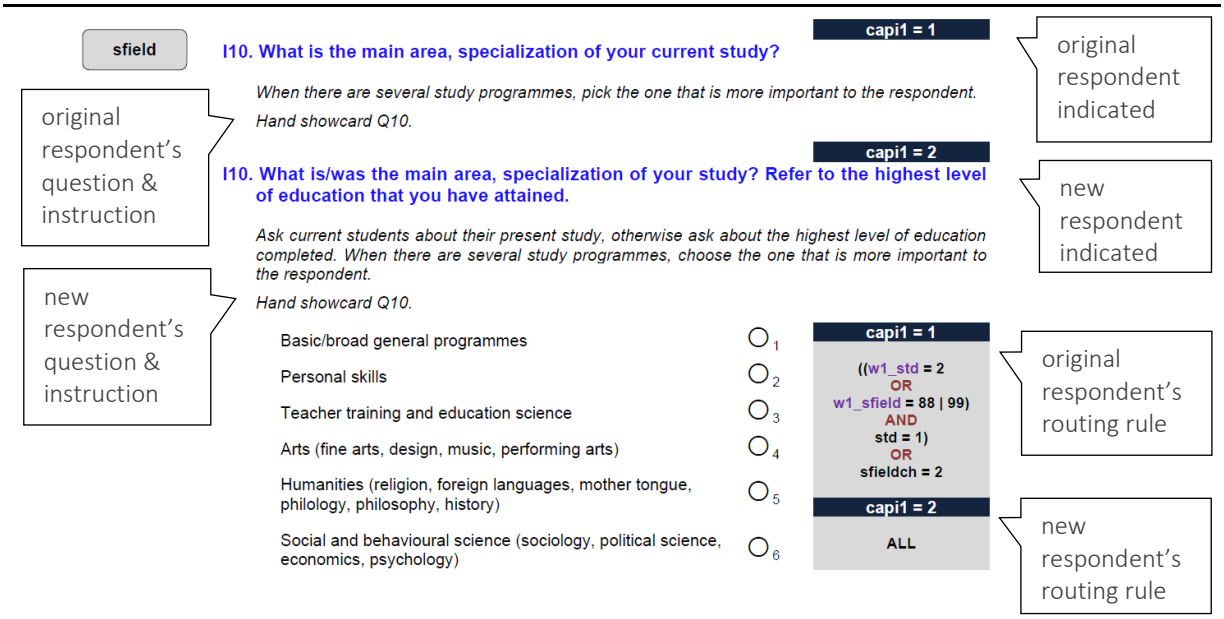
The questionnaires also specify any interviewer alerts in case certain unlikely values have been entered, and the procedure for subsequent imputation of values for dependent interviewing variables.

Figure 3: Other constituent parts of the CAPI questionnaires published



From Wave 2, the CAPI individual questionnaire includes a routing condition based on whether the respondent completed the same questionnaire in the previous wave/waves. This condition is marked separately from the basic routing rule, namely in the question header (white text on dark-blue background). When the question is worded differently for original and new respondents, then both wordings are given under the routing specification. Any basic routing conditions that are not the same for old and new respondents are also divided in this way. *Figure 4* illustrates the situation on the *sfield* variable in Wave 2.

Figure 4: Differentiating question parameters for original and new respondents in the CAPI individual questionnaire



The self-completion instruments are published in the form distributed to respondents; only the dataset's variable codes were added. In some cases, adding the codes required minor adjustments to the layout and formatting of the questionnaire (indentation, font size).

The fieldwork agencies reformatted the question coding symbols in the self-completion instruments in line with their data entry procedures. MEDIAN scanned the questionnaires and answers had to be checked in dedicated boxes (see questionnaire example in *Figure 5*). Since STEM/MARK entered the data manually, the questions in its questionnaire were accompanied by larger-size numeric response codes for circling, instead of the checkboxes.

Both agencies used the same format of printed diaries. Like in the case of self-completion questionnaires, variable codes were added to their published versions.

Figure 5: Comparing the original and published forms of the self-completion questionnaire

O.1 All things considered, how satisfied are you with your life as a whole nowadays?

extremely dissatisfied	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9	☐ 10	extremely satisfied

O.2 How much do you agree or disagree with the following statements? *(cross one answer in each row):*

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
I often feel stressed and under time pressure.	☐1	☐2	☐3	☐4	☐5
I regularly get enough sleep.	☐1	☐2	☐3	☐4	☐5
I would like to have more time for myself.	☐1	☐2	☐3	☐4	☐5
I would like to have more time for my family.	☐1	☐2	☐3	☐4	☐5
I would like to have more time for advancing my career.	☐1	☐2	☐3	☐4	☐5

lfsat **O.1 All things considered, how satisfied are you with your life as a whole nowadays?**

extremely dissatisfied	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐ 8	☐ 9	☐ 10	extremely satisfied

O.2 How much do you agree or disagree with the following statements? *(cross one answer in each row):*

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
timestre I often feel stressed and under time pressure.	☐1	☐2	☐3	☐4	☐5
timesle I regularly get enough sleep.	☐1	☐2	☐3	☐4	☐5
timemys I would like to have more time for myself.	☐1	☐2	☐3	☐4	☐5
timefam I would like to have more time for my family.	☐1	☐2	☐3	☐4	☐5
timecar I would like to have more time for advancing my career.	☐1	☐2	☐3	☐4	☐5

4.5. CAWI mutations of data collection instruments

The CAWI mutations of questionnaires were prepared in maximum conformity with the original CAPI and self-completion versions, although the specifics of web-based interviewing dictated adjustments in several cases.

In the CAWI versions of CAPI questionnaires, the codes for invalid answers "don't know, can't say" and "refused" were visible to the respondents, whereas those options were not prompted in face-to-face interviewing. For the questions *partycl* and *partywv,* the respondents stated the names of political parties spontaneously, and the interviewers picked the corresponding answers from a list; the list was prompted in the CAWI mutation. In Wave 2, a large majority of the cognitive testing section was dropped from the CAWI mutation of individual CAPI.

In the CAWI versions of self-completion instruments, the respondents went through one question on its own screen at a time (they were, however, able to return to previous questions). The CAWI diaries were divided into four screens with one-fourth of the day each. In both types of instruments (self-completed questionnaires and diaries), the CAWI respondents had to pick an answer to every question, whereas this check could not be implemented for the pen-and-paper versions.

4.6. Cross-wave coding consistency

For an overwhelming majority of questions repeated in several waves of the survey, the same range of response options and codes were preserved. Data files from each wave contain the full range of value labels used across all waves (including inapplicable and error/missing codes), even though some of these labels may not be applicable to all waves (e.g. the variable *rstat* in the W1 dataset contains a label for code 5 which was introduced only later in Wave 2).

For a few variables, the answer options were reworded between waves while their meaning remained unchanged (such as in the variables *asame* or *rel*). The data files contain labels corresponding with the wording used in the respective wave.

The list of political parties changed for the questions *partycl*, *partyvt* and *partywv*, and therefore their numbering in the questionnaires used to collect the data. In the data files, party codes were harmonized across waves, and these harmonized codes were included in the published questionnaires, but with the original order of parties. The routing rule for the variable *hclose* as well as showcards contain the original party numeric codes. Party names mentioned in semi-open questions (*partyclo, partywvo*, etc.) were coded into corresponding party codes, even though such code was introduced only in a later wave. The data thus contains also codes of such parties that were not included in response options that year.

Before Wave 2, the answer options for the *refam*, *rehous* and *reelse* questions were reworded based on experience with the categories used in Wave 1. The options prompted differ in terms of both content and numeric codes.

In variables *psapet* to *psaint*, response options were adjusted for Wave 3 to refer to the happenings in the past year, whereas options offered in Wave 2 related to the past in general. The same codes in Wave 2 consequently have different meanings that those in the following waves.

Table 18: Variables with coding differences

asame	1–2	answers reworded between Waves 2 and 3, meaning unchanged
rstat	1–4	answers reworded between Waves 1 and 2, meaning unchanged
rstat	5	code added to Wave 2
new	1–2	answers reworded between Waves 2 and 3, meaning unchanged
capistat	1–3	content of derived variable values changed between Waves 3 and 4
mstat	2	answers reworded between Waves 1 and 2, meaning unchanged
educ	0	answers reworded between Waves 1 and 2, meaning unchanged
rel(*x*)	3	answers reworded between Waves 1 and 2, meaning unchanged
rel(*x*)	5	answers reworded between Waves 2 and 3, meaning unchanged
rel(*x*)	8	answers reworded between Waves 2 and 3, meaning unchanged
rel(*x*)	10	answers reworded between Waves 2 and 3, meaning unchanged
rel(*x*)	18	answers reworded between Waves 2 and 3, meaning unchanged
dwelob	1–8	answers reworded between Waves 1 and 2, meaning unchanged
unemre	9	closed option changed to open between waves 1 and 2
partycl	11	closed option changed to open between waves 1 and 2
partycl	12	closed option changed to open between waves 3 and 4
partywv	11	closed option changed to open between waves 1 and 2
partywv	12	closed option changed to open between waves 3 and 4
refam	1–4	content of options changed between Waves 1 and 2
refam	5–10	code added to Wave 2
rehous	2	answers reworded between Waves 1 and 2, meaning unchanged
rehous	4	content of option changed between Waves 1 and 2
rehous	5	code added to Wave 2
reelse	4	content of option changed between Waves 1 and 2
reelse	5–9	code added to Wave 2
mdwelob	1–9	answers reworded between Waves 1 and 2, meaning unchanged
psabou-psaral	1–3	content of options changed between Waves 2 and 3

4.7. Looped variables

Selected questions in the household and individual CAPI were repeated for all relevant individuals, events etc. For these variables, the routing rule area in the questionnaire contained a looping criterion.

A question was repeated as many times as prescribed by the *loop* criterion and any other routing conditions. When the criterion was set in line with the *hhsize* variable, then the maximum number of loops was 12 in Wave 1 and 10 in Wave 2. For other numerical variables, the maximum number of loops was indicated in the questionnaire. The order of a variable in the loop was indicated by the suffix (x) in its code (without the brackets).

In individual CAPI data, the information from looped questions was recorded in separate variables coded with a suffix indicating the loop order (e.g., *eduex1*, *eduex2*; generally referred to as (*x*) in the questionnaire). In contrast, answers to looped questions in the CAPI household questionnaires were recorded in a single variable attached to the household member they referred to.

Looping was typically done within variables: first the question was repeated as many times as the number of relevant cases, and only then did the interview move on to the next question. Only in the case of employment history variables (Wave 2), all questions pertaining to the same job, period of unemployment etc. were asked first and only then did the interview continue with the next time period. The variables in the dataset are ordered in the same way the questions were asked during the interview.

Household grid variables in the introductory part of the household CAPI represent a special case: the interviewer/respondent was able to complete information both vertically (one person at a time) and horizontally (one variable at a time).

4.8. Dependent interviewing

In Waves 2 to 4, dependent interviewing was employed in the household and individual CAPI. One's trajectory through the questionnaire or the wording of one's questions were contingent upon the information obtained in previous waves.

Routing of many questions in both instruments is based on the values of previous waves' variables. The household CAPI always started by showing the previous wave's address and inquiring whether the household still resided at the same address. This variable, *asame*, then played a key role in the household interviewing trajectory. Households that remained on the same address were not asked about stable housing characteristics. The individual CAPI trajectory was fundamentally shaped by whether the person had completed the CAPI individual questionnaire in at least one of the preceding waves. Those who had (so-called original respondents) were no longer asked about their family background or employment history; such questions were only asked of new respondents. In contrast, to prevent excessive interview duration, new respondents in Wave 2 were not asked all employment history questions.

Key questions about individuals and households (list of household members and their basic sociodemographics, housing tenure, employment status, etc.) were not asked again. Instead, respondents were merely asked to confirm the validity of last wave's data, or they were reminded of

the information before answering a question (so-called proactive dependent interviewing). Reactive dependent interviewing, where the respondent is reminded of the existing value in case of a discrepancy, was not employed.[4]

Dependent interviewing in Wave 2 relied on the values of Wave 1 variables. In the question routing and wording in the questionnaires, these variables are marked with the prefix *w1* before an underscore (e.g., *w1_name*) and using purple font colour (see *Figure 3*). Dependent interviewing in Waves 3 and 4 relied on auxiliary variables constructed from pervious waves' information. Such auxiliary variables are shown in Wave 3 and 4 questionnaires with the prefix *o* before an underscore (e.g., *o_name*) and using purple font colour. When a previous wave's answer was included in the wording of a question, this is indicated by the variable code in angle brackets such as *<w1_htnr>* (see *Figure 3*). When question wording was modified depending on a previous wave's value, this is indicated using red font colour.

4.9. Cognitive testing (Wave 2)

The second wave of the survey included several tasks to test the respondents' cognitive ability. These cognitive tests were included in the individual CAPI and, as such, they were only administered to this instrument's respondents. Since they could not be administered without an interviewer, they were excluded from the CAWI mutation of the instrument. CAWI respondents only answered the introductory question on an overall evaluation of their memory (the *selfmem* variable), and their other cognitive testing variables were coded as 97/999,997.

The tests were adopted from Wave Three of the UK Household Longitudinal Study (Understanding Society).[5] An effort was made to word the questions and interviewer instructions as close to this source as possible. The biggest changes were made to the numeracy exercises, in which numerical information was adapted to the Czech context. At the same time, one's trajectory through these exercises was not dependent on their score in the preceding numeracy exercise.

A special section of the interviewer training was dedicated to the cognitive testing. Interviewers were instructed to avoid presenting the exercises as intelligence tests, and instead to frame them as a "refreshment" during the interviewing session. They were strictly dissuaded from showing the screen to the respondents in this part of the questionnaire, giving them advice, or assisting them in any way. Even though interviewers strived to organize the testing privately, in the absence of other household members, this was not always possible to arrange.

The episodic memory test started by listening to a recording of ten words. Subsequently, the respondent was invited to repeat, in any order, as many words he/she could recall as possible. The time to recall the words was not limited. When the recording could not be started due to technical difficulties, the interviewer read the words out at a pace comparable to the recording. Each version was randomly

[4] For more on types of dependent interviewing, see, for instance, Jäckle, A. 2009. Dependent Interviewing: A Framework and Application to Current Research. Pp. 93–11 in P. Lynn (ed.) *Methodology of Longitudinal Surveys.* Chichester: Wiley.

[5] McFall, S. 2013. *Understanding Society: UK Household Longitudinal Study: Cognitive Ability Measures.* Colchester: University of Essex, Institute for Social and Economic Research.

administered to one half of the respondents. The *wordver* variable indicates which set of words was used with the respondent.

In the verbal fluency test, the interviewer first read the instructions and started a 60 seconds countdown timer, during which the respondent was to say as many names of animal species as possible. At the same time, the interviewer was writing the names down (using abbreviations or tally marks instead of whole words if necessary) on a dedicated sheet with notes on instructing the respondent and evaluating the exercise. The exercise was recorded to allow for verification of the number of animals named. In the course of data cleaning, the fieldwork agencies compared the interviewer-input number with the tally sheets and recordings, correcting any discrepancies.

The numeracy test comprised of four exercises focusing on practical numeracy. After the interviewer read the instructions, the respondents were to formulate the answers independently. The interviewer then recorded whether the respondent's answer complied with the correct solution ("correct answer") or not ("incorrect answer").

In the cognitive exercises (except the memory self-assessment, *selfmem*), the invalid answer options "don't know, can't say" and "refused" were not available to the interviewer.

4.10. Employment history (Wave 2)

In Wave 2, the CAPI individual questionnaire included a section mapping the respondent's complete labour market trajectory. However, these questions were only administered to respondents who had completed the individual CAPI in Wave 1. Since new individual CAPI respondents had to answer questions on their educational trajectories or family background, which were not applicable to the original respondents, their questionnaires were more time-consuming.

In accounting for their employment history, the respondents did not proceed chronologically. Instead, they started by stating the number of periods of each kind (e.g., number of jobs lasting more than one year) and then characterized the different time periods by answering additional questions.

To facilitate respondents' orientation in the questions and help them better recall the different time periods and place them on the timeline, the interviewers could make use of an employment history timeline template. The timeline from 1955 to 2016 was printed on the first page of the document, with separate sections for recording periods of (a) compulsory military service, (b) employment, (c) unemployment and (d) maternity/parental leave, plus fields for recording the number of such periods. The second page showed two different examples of employment histories with completed timelines.

The use of timelines to support questionnaire completion was optional and, in contrast to the animal naming sheet, self-completion questionnaires and diaries, the timelines were not delivered by the interviewers to their agencies.

5. Dataset content and structure

The datasets documented herein contain variables from the CAPI questionnaires and the adult and child versions of the self-completion questionnaires. The diary variables are stored in separate datasets with their own documentation. A separate MS Excel file provides a complete overview of variables in the data.

5.1. Variable organization

The variables from the different instruments are organized in line with *Table 19*. Variables originating from several different instruments are grouped with variables of the lower-order instrument. As an exception, the *health* and *hamper* variables coming from the proxy CAPI, and the adult self-completion questionnaires are grouped with variables from the adult self-completion. For individuals who completed an adult self-administered questionnaire and, at the same time, had a CAPI proxy questionnaire completed for them, the *health* and *hamper* variables take the values from the adult self-completion (except when that question was not answered in the adult self-completion).

The variable labels contain a letter indicating the questionnaire and the question number within that instrument. This does not apply to IDs, derived variables and interviewing context variables.

Table 19: Organization of groups of variables in the dataset

Household/person ID	
CAPI household questionnaire	H + question number in variable label
CAPI individual and proxy questionnaire	I, P + question number in variable label
Adult self-completion	A + question number in variable label
Child self-completion, age 15–17	Y + question number in variable label (YC from Wave 3)
Child self-completion, age 10–14	C + question number in variable label (YC from Wave 3)
Stratification status	
Household place of residence	
Interviewing characteristics	
Interviewing times	
Questionnaire completion within household	
Instrument completion indicators and weights	

5.2. Dataset structure

The cases in the datasets represent members of the surveyed households (13,077 in Wave 1, 10,773 in Wave 2 etc.) Household-level and individual-level data from the household CAPI are available for every member. The values of household-level variables are equal for all household members. As an exception, for individuals who were no longer members of the household at the time of interviewing, most household-level variables were coded as invalid (97/999,997; for details on leavers, see Section 7, *Household composition change variables*).

The values of variables from other instruments (individual and proxy CAPI, self-completions) are only available for respondents who completed them, else the value is system-missing. *Table 20* illustrates the structure of the dataset.

Table 20: Dataset structure

hid	pid	Household CAPI, household level	Household CAPI, individual level	Individual/proxy CAPI	Adult self-completion	Child self-completion
1	11	1	11	11	11	.
2	21	2	21	21	.	.
2	22	2	22	22	22	.
2	23	2	23	.	.	23
3	31	3	31	31	31	.
3	32	3	32	32	32	.
4	41	4	41	41	.	.
5	51	5	51	51	51	.
5	52	5	52	.	.	52

5.3. Variables not included in the dataset

The datasets do not contain all questions/variables from the data collection instruments. Excluded were variables with more detailed information about the respondent (e.g., first name that served to distinguish between household members during interviewing; the description of respondent's job; address of the previous place of residence) and results of checks in the course of interviewing (e.g., *namech*). *Table 21* lists such excluded variables.

Also excluded were variables related to the household's 11th and 12th members (the households surveyed had a maximum of 10 members; from Wave 2, the questionnaire limit was lowered to 10 household members) and variables from partnership history included in adult self-completion in Wave 2 pertaining to partners of unused order (7th to 10th).

Table 21: Variables not included in the dataset			W1	W2	W3	W4
achange	reason for address change	CAPI household questionnaire	-	-	1	1
name	given name (or alias)	CAPI household questionnaire	1	1	1	1
namech	check to verify entry: name	CAPI individual questionnaire	1	1	1	1
sexch	check to verify entry: sex	CAPI individual questionnaire	1	1	1	1
birthmch	check to verify entry: month of birth	CAPI individual questionnaire	1	1	1	1
birthych	check to verify entry: year of birth	CAPI individual questionnaire	1	1	1	1
educch	check to verify entry: education	CAPI individual questionnaire	1	1	1	1
stdych	check to verify entry: study status	CAPI individual questionnaire	1	1	1	1
estatch	check to verify entry: economic status	CAPI individual questionnaire	1	1	1	1
mstatch	check to verify entry: marital status	CAPI individual questionnaire	1	1	1	1
sname(*x*)	name of school presently attended	CAPI individual and proxy questionnaire	1	1	1	1
sstr(*x*)	street of school presently attended	CAPI individual and proxy questionnaire	1	1	1	1
smun(*x*)	municipality of school presently attended	CAPI individual and proxy questionnaire	1	1	1	1
sprog(*x*)	programme presently studied	CAPI individual and proxy questionnaire	1	1	1	1
swrkps	description of job in between waves	CAPI individual questionnaire	1	1	-	1
nwrkps	future job description	CAPI individual questionnaire	1	1	-	1
wrkps	respondent's job description	CAPI individual and proxy questionnaire	1	1	1	1
mstatcor	whether marital status same as during previous interview	CAPI individual questionnaire	1	1	-	1
mcorch	marital status during previous interview	CAPI individual questionnaire	1	1	-	1
pwrkps	respondent's partner's job description	CAPI individual questionnaire	1	1	-	1
liwhyo	other reason why respondent did not live predominantly with both biological parents till 16 years of age	CAPI individual questionnaire	1	1	1	1
fwrkps	respondent's father's job description when respondent was 16 years old	CAPI individual and proxy questionnaire	1	1	1	1
mwrkps	respondent's mother's job description when respondent was 16 years old	CAPI individual and proxy questionnaire	1	1	1	1
mstreet	street of previous place of residence	CAPI individual questionnaire	1	1	1	1
mcounty	county of previous place of residence	CAPI individual questionnaire	1	1	1	1
mmunic	municipality of previous place of residence	CAPI individual questionnaire	1	1	1	1
mpsc	ZIP code of previous place of residence	CAPI individual questionnaire	1	1	1	1
hwrkps(*x*)	past job description	CAPI individual questionnaire	-	-	-	1

papiname	given name indicated in the self-completion	Adult and child self-completion	1	1	1	1
papiage	age indicated in the self-completion	Adult and child self-completion	1	1	1	1
papisex	sex indicated in the self-completion	Adult and child self-completion	1	1	1	1
jobas	aspirational future job description	Child self-completion	1	1	1	1
jobex	expected future job description	Child self-completion	1	1	1	1

6. Identification variables

Households and individuals were assigned numbers allowing for unique identification of every case. Household and person IDs are wave-specific, and only corresponding identifiers can be used for linking datasets from multiple waves. The ID's prefix indicates to which wave it belongs (e.g., *w1_hid*). To link individuals between Waves 1 and 2, it is necessary to combine Wave 1 *w1_pid* with Wave 2 *w1_pid* (rather than Wave 2 *w2_pid*); data from other waves can be combined accordingly.

Household members in each dataset are assigned IDs from all preceding waves, but not from following ones. ID values are unique within a given wave only. For example, Wave 2 dataset contains duplicate values of *w1_pid* in cases of split households when more than one secondary household was surveyed. For more information about duplicates, and distinguishing between primary and secondary occurrence of a case in the wave, refer to section 7, *Household composition change variables*.

Households are differentiated by unique identifiers, *hid*, shared by all their members. In Wave 2, for the indicator of Wave 1 *hid* (*w1_hid*), the digit 0 on the second position is replaced by the value of *asame* variable. When the original household split and all its parts remained at the same address, the household's order (set arbitrarily) appears on the second position of *hid.* The household IDs in Waves 3 and 4 are constructed in the same way: *w3_hid* is derived from *w2_hid* in which the digit 0 on the third position is replaced by the Wave 3 value of *asame,* and *w4_hid* is derived accordingly from *w3_hid* and Wave 4 *asame.* If the *asame* value from a previous wave was subsequently corrected, the respective *hid* identifier was kept intact. In some cases, the *asame* value, therefore, does not match the corresponding digit in the identifier.

The unique person ID (*pid*) in each wave is a combination of household ID (*hid*) and the person's order in that household (*pno*). The unique person ID (*pid*) defines the respondent of the individual CAPI and self-completion questionnaires, or the person for whom a proxy CAPI was completed. The ID of the proxy CAPI respondent (the person who completed that questionnaire, not the one to which their answers relate) is defined by the *pidpro* variable, and the household CAPI respondent's ID is defined by *pidhr*.

Table 22: Household/individual identification variables

w(*x*)_hid	unique household ID in Wave (x)
w(*x*)_pid	unique personal ID in Wave (x)
pno	person's order in the household
pidhr	ID of the person who completed the household questionnaire
pidpro	ID of the person who completed the proxy questionnaire

Every person in the dataset is assigned IDs of other household members with a defined relation to that person (*Table 23*).

Table 23: Other household members' identification

pidp	ID of household member who is the person's spouse (*rel* = 1), cohabiting partner (*rel* = 2) or registered same-sex partner (*rel* = 3)
pidch(*x*)	ID of household member whose *pno* = (x) and who is the person's biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) child
pidf *)	ID of the man (*sex* = 1) in the household who is the person's biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster father (*rel* = 12)
pidm *)	ID of the woman (*sex* = 2) in the household who is the person's biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster mother (*rel* = 12)
pidgf	ID of the man (*sex* = 1) in the household who is the person's grandfather (*rel* = 21)
pidgm	ID of the woman (*sex* = 2) in the household who is the person's grandmother (*rel* = 21)

*) Persons who have in their household a step parent of the same sex as their biological parent are also assigned an ID in the variables representing the opposite sex to that of the biological parent.

7. Household composition change variables

Changes in the composition of households between waves are reflected in several variables that were either asked as questions or generated from other variables (*Table 24*). These variables are included in the datasets from Wave 2 on (except *rstat* which was already included in Wave 1).

In Waves 2 to 4, the last wave's household grid (first names, sex, age etc.) was presented in the introductory section of the CAPI questionnaire. Interviewers then invited the respondents to indicate in the *rstat* variable (household member status) for each person whether they still belonged to the household (codes 1–4) or this was no longer the case for various reasons (death, relocation etc.). Persons with *rstat* equal to 5 (no longer member of my household) are referred to as leavers. In Waves 3 and 4, previous waves' leavers were also included in the household grid, except for those who had deceased (*leftre* = 1). The current value of *rstat* for previous waves' leavers indicates any rejoiners (leavers who returned to the household). The interviewers and respondents added to the household grid all joiners who had not belonged to the household in any previous wave.

The *hhmem(*x*)* variable indicates a person's household membership in the wave that corresponds to the variable's numerical suffix. The variable is never available for the current year (e.g., *hhmem2* in Wave 2 data) because it is the *rstat* variable that indicates the respondent status in the current wave. The *hhmem(*x*)* variable equals 1 (yes) for persons who were indicated as currently residing or temporarily absent members (*rstat* = 1/2/3/4/88/99). The value 2 (no) indicates two types of persons: either leavers (individuals who had left the household before the current wave, *rstat* = 5) or future joiners (individuals who do not belong to the household yet but will be born or move into the household in one of the following waves). Such individuals are also identified by code 97 in their person ID, *pid*, of the relevant wave.

All leavers (including deceased ones) are included in the dataset in addition to current household members. As a result, duplicates arise in the course of surveying both parts of split households. Their members are included twice in the data: first as leavers from their former households and second as current members of their new households. Current wave's personal ID (*pid*) is always unique (since it is derived from household ID, *hid*), but previous waves' IDs are not. The *primary* variable serves to indicate "primary presence" (i.e. the household in which one is currently a member). The *primary* variable always relates to the current wave only. A secondary occurrence of a case in a given wave may thus become primary again in the next wave.

The *new* variable indicates the presence of at least one joiner, compared to the previous wave. The related *newn* variable indicates the number of joiners. From Wave 3, individuals who had previously left the household and then returned to it (rejoiners) were also included among joiners. Similarly, the *leavers* variable indicates the number of leavers. In Waves 3 and 4, however, only individuals who had left the household between the immediately preceding and current waves were included among leavers. In contrast to the above-mentioned individual-level variables, the variables *new*, *newn* and *leavers* relate to the level of household.

The household-level variable, *hchange,* indicates whether or not household composition has changed. It has changed if at least one member has joined or left the household.

Table 24: Variables representing household joiners and leavers

rstat	from Wave 2, categories 1 to 4 representing a household member's residential status were supplemented with category 5 to identify a leaver
hhmem(*x*)	household membership in Wave (x)
primary	person occurres in the sample primarily (member of unsplit household, current member of split household) or secondarily (leaver from split household). The value refers to the current wave only (it is therefore possible that the secondary occurrence changes back to primary).
new	presence of joiners (including returning leavers = rejoiners)
newn	number of joiners (including returning leavers = rejoiners)
leavers	number of leavers (left household between current and immediately preceding wave)
hchange	indicates change in household composition (i.e. existence of joiners or leavers)

In the CAPI variables, leavers are assigned valid values (as long as other routing conditions are met) of the variables *w(*x*)_hid, w(*x*)_pid*, *pno*, *asame*, *new*, *newn*, *rstat*, *sex*, *birthm*, *birthy*, *age*, *agecat*, *primary*, *hhmem(*x*)*, *capi(*x*)*, *hhsize*, *leavers*, *leftre*, *leftreo*, *leftm*, *lefty, lcounty, lcont, lmstat, lwith, leduc* and *hchange*. Other variables from this questionnaire hold inapplicable codes for leavers (97/999,997). In the variables from other instruments, leavers have system-missing values. However, values of interviewing context variables are assigned to them. In variables indicating instrument completion (*hh_hh*, *hh_indi* etc.), leavers have zero values, and no weights are computed for them.

The numbers of joiners and leavers in each wave are shown in *Table 25*. The *households* column indicates how many households have at least one member of each type. Number of leavers is indicated both including and excluding duplicate individuals (i.e. members of split households who appear in the dataset twice because more than one secondary household has been surveyed).

Table 25: Numbers of joiners and leavers

	W2		W3		W4	
	households	individuals	households	individuals	households	individuals
joiners	235	258	234	278	175	204
leavers including duplicate individuals	224	282	200	240	160	206
leavers excluding duplicate individuals	192	231	178	205	144	166
duplicate individuals	32	51	22	35	17	40

8. Dependent interviewing variables

To make answering questions easier for respondents, render the interviewing process more efficient and reduce errors, the household and CAPI individual questionnaires in Waves 2 to 4 were adjusted to take into account the previous waves' characteristics of individuals and households (so-called dependent interviewing).

Since the respondent's route through individual CAPI depended on his/her previously having completed that type of instrument, a set of indicators to distinguish between original and new respondents was prepared. The *capi(*x*)* variable indicates whether an individual had completed the CAPI individual questionnaire in the year indicated by the variable's numerical suffix. Each dataset only contains such indications related to previous waves (the *indi* variable identifies current wave's individual CAPI respondents). The variable *capitat* in Waves 3 and 4 indicates various combinations of individual CAPI completion during previous waves. To simplify the formulation of routing conditions, the *capistat* variable was dichotomized into *capi,* which indicates whether the respondent had completed at least one individual CAPI in the past.

The CAPI household and individual questionnaires used selected previous waves' variables in the following ways (*Table 26* summarizes the types of use of the different variables):

1 An answer was pre-completed with the previous wave's value. If a change had occurred or a wrong value had been recorded, the respondent was to select a new answer;

2 Respondent was reminded of a previous wave's answer in the wording of the question. Respondent stated whether or not the situation had changed;

3 Respondent was reminded of a previous wave's answer in the wording of the question. Respondent specified what the current situation was;

4 The value was used to define routing rules, not presented to the respondent.

When the previous answer was presented in inquiring about the current situation (types 2 and 3 in the overview above), values of target variables were imputed during cleaning in line with the following rules. When the respondent's answer to a comparison question (e.g., *wrkpsch*) indicated no change (type 2), the previous value was imputed in the target variable (e.g., *wrkps*). For example, when the respondent's answer to *wrkpsch* in Wave 2 indicated that he/she was still in the same job as in the previous wave, then his/her Wave 1 job was imputed in the *wrkps* variable. For questions where the previous value was presented and then the respondent specified the current situation, not just compared it to the former (type 3), the current situation indicated by the comparative variable (e.g., *svisech*) was imputed in the target variable (e.g., *svise*). For example, when the respondent answered the comparative question *svisech* in Wave 2, this answer was imputed in the target variable *svise.* The details of imputation are specified in the given variable's area of the questionnaire.

Note that imputation was not performed for questions that were not administered to original respondents of individual CAPI at all (e.g., *eduprim*, *edusec*). For these respondents, the values of all

such variables are only available in the dataset of the wave in which the respondent first completed his/her individual CAPI.

Whereas dependent interviewing in Wave 2 simply relied on Wave 1 values, the subsequent waves had to take into account that the required values for the household or individual may not be available in the immediately preceding wave. Especially in the case of values from individual CAPI, it was necessary to consult the last wave in which the respondent had completed this instrument, not the previous wave. In Waves 3 and 4, auxiliary variables were prepared for dependent interviewing purposes which indicated the last relevant wave's value for the given household or individual.

Values of the previous wave (for household CAPI variables) or of the last wave in which the respondent had completed their individual CAPI (for individual CAPI variables) were typically imputed in these auxiliary variables. In a few cases, however, the values of earlier waves were imputed. The values of the last wave in which a person was a member of the household (i.e. was not a leaver, *rstat* = 5) were imputed in the auxiliary variables *o_mstat*, *o_educ* and *o_rel(*x*)*. This meant the previous wave for most individuals, but the wave before leaving the household for leavers. The auxiliary variables *o_dwelob*, *o_faliv*, *o_maliv* and *o_fowny* could not work with the value of the previous wave (for households) or of the last completed individual CAPI because ongoing imputation of values did not take place in the different waves (see above). Therefore, the answer from the last wave in which the question had been asked was imputed in the auxiliary variables.

During preparation of previous waves' data for use in dependent interviewing, cleaning was performed on the string variables *sname1* (standardization of school names including imputation of street and city from the *sstr1* and *smun1* variables), *sprog1* and *wrkps* (correction of spelling mistakes and typos).

Table 26: Dependent interviewing variables

(target) variable	type of use	comparison variable	
capi(*x*)		-	indicates whether household member completed CAPI individual questionnaire in Wave (x)
capistat	-	-	indicates completion of CAPI individual questionnaire in previous years
capi	-	-	indicates completion of CAPI individual questionnaire in at least one of the previous years
name	1	-	given name of household member
sex	1	-	sex of household member
birthm	1	-	month of birth of household member
birthy	1	-	year of birth of household member
mstat	4	-	marital status of household member
educ	1	-	educational attainment of household member
rel(*x*)	1	-	relation of household member to persons *pno* = x
htnr	2	htnrch	housing tenure
dwelob	4	-	how was the current accommodation obtained
sfield	4	-	field of study
stype	4	-	school attended/type of programme studied
sname1 *)	2	snamech1	name of school presently attended
sprog1 *)	2	sprog1ch	programme presently studied
wrkps *)	2	wrkpsch	respondent's job description
isco08	2	wrkpsch	respondent's job description coded in accordance with the CZ-ISCO classification
svise	3	svisech	whether respondent supervises other employees
suborch	2	subor	number of respondent's subordinates
empl	3	emplch	whether respondent has employees
emplsch	2	emplsch	number of respondent's employees
fsizech	2	fsizech	number of employees of respondent's employer
contr	2	contrch	type of employment contract
faliv	4	-	person's father is alive
maliv	4	-	person's mother is alive
fowny	4	-	year when person first became a property owner

*) Excluded from the dataset in the process of anonymization.

9. Derived variables

The dataset contains several variables whose values are derived from other variables. Some of these were generated in the course of interviewing and for routing purposes, others were constructed ex post.

Household-level derived variables have equal values for all household members. Individual-level derived variables based on household CAPI variables hold values for all household members. As an exception, leavers were only assigned valid values for selected variables of both these types (for details, see Section 7, *Household composition change variables*).

Individual-level derived variables constructed from individual or proxy CAPI variables have values for those respondents who had completed the instrument in the given year (with the variable *sumchild* being an exception, as explained in *Table 29*).

Table 27: Household-level derived variables (CAPI household questionnaire)

hhsize	number of persons in the household
nborn	number of persons in the household aged under 1 year (*age* < 1)
ch0to5	number of persons in the household aged up to 5 years (*age* ≤ 5)
ch6to11	number of persons in the household aged from 6 up to 11 years (6 ≤ *age* ≤ 11)
ch10to17	number of persons in the household aged from 10 up to 17 years (10 ≤ *age* ≤ 17)
ch0to17	number of persons in the household aged up to 17 years (*age* ≤ 17)
p0to26	number of persons in the household aged up to 26 years (*age* ≤ 26)
adults	number of adult persons in the household (*age* ≥ 18)
new	presence of joiners in the household
newn	number of joiners in the household (*hhmem(x)* = 2 / *w(x)_pid* = 97)
leavers	number of leavers (*rstat* = 5)
hchange	indicates change in household composition (i.e. existence of joiners or leavers)
nown	whether household became housing property owner since last wave

From Wave 2, the CAPI household questionnaire contained the variable *#hhsize* indicating the total number of individuals linked to the household, i.e. including leavers (but not deceased ones). This auxiliary variable was only used in interviews to account for leavers, and therefore, it is not included in the dataset.

Several indicators of household structure were derived from household-level variables, namely the *hhtype* variable and the set of variables, *hhtype2* to *hhtype6* (*Tables 28* and *29*). The variables *hhtype2* to *hhtype6* indicate the presence of defined groups of individuals in the household irrespective of the presence of any other individuals, whereas *hhtype* indicates exclusive presence of the groups defined. For example, if a married couple lived with the husband's mother (wife's mother-in-law), their *hhtype* would equal 7 (other) and their *hhtype2* and *hhtype3* would equal 1 (yes).

Table 28: Indicators of presence of groups of persons in the household

hhtype2	whether household contains (at least one) group of persons meeting the following conditions: two members form a married (*rel* = 1), cohabiting (*rel* = 2) or same-sex partnership couple (*rel* = 3)
hhtype3	whether household contains (at least one) group of persons meeting the following conditions: an individual who does not have his/her spouse (*rel* ≠ 1), cohabiting partner (*rel* ≠ 2) or registered same-sex partner (*rel* ≠ 3) in the household is the biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster parent (*rel* = 12) of at least one more member, none of whom is younger than 18 years (*age* ≥ 18)
hhtype4	whether household contains (at least one) group of persons meeting the following conditions: an individual who does not have his/her spouse (*rel* ≠ 1), cohabiting partner (*rel* ≠ 2) or registered same-sex partner (*rel* ≠ 3) in the household is the biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster parent (*rel* = 12) of at least one more member, of whom at least one is younger than 18 years (*age* < 18)
hhtype5	whether household contains (at least one) group of persons meeting the following conditions: two members who form a married (*rel* = 1), cohabiting (*rel* = 2) or same-sex partnership couple (*rel* = 3) are the biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster parents (*rel* = 12) of at least one more member, none of whom is younger than 18 years (*age* ≥ 18)
hhtype6	whether household contains (at least one) group of persons meeting the following conditions: two members who form a married (*rel* = 1), cohabiting (*rel* = 2) or same-sex partnership couple (*rel* = 3) are the biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster parents (*rel* = 12) of at least one more member, of whom at least one is younger than 18 years (*age* < 18)

Table 29: Household types by structure

1	one-person household	• household has one member (*hhsize* = 1)
2	married, same-sex partnership or cohabiting couple	• household has two members (*hhsize* = 2), • household members form a married (*rel* = 1), cohabiting (*rel* = 2) or same-sex partnership couple (*rel* = 3)
3	single parent with adult child/children	• household has at least two members (*hhsize* ≥ 2), • one household member is the biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster (*rel* = 12) parent of all others, • other members are biological (*rel* = 13), step (*rel* = 14, *rel* = 15), adoptive (*rel* = 16) or foster (*rel* = 17) siblings to one another, • none of the other members (children) is younger than 18 years (*age* ≥ 18)
4	single parent with at least one child under 18 years	• household has at least two members (*hhsize* ≥ 2), • one household member is the biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster (*rel* = 12) parent of all others, • other members are biological (*rel* = 13), step (*rel* = 14, *rel* = 15), adoptive (*rel* = 16) or foster (*rel* = 17) siblings to one another, • at least one of the other members (children) is younger than 18 years (*age* < 18)
5	parents with adult child/children	• household has at least three members (*hhsize* ≥ 3), • two members form a married (*rel* = 1), cohabiting (*rel* = 2) or same-sex partnership couple (*rel* = 3), • the couple consists of the biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster (*rel* = 12) parents of all others, • the other members are biological (*rel* = 13), step (*rel* = 14, *rel* = 15), adoptive (*rel* = 16) or foster (*rel* = 17) siblings to one another, • none of the other members (children) is younger than 18 years (*age* ≥ 18)
6	parents with at least one child under 18 years	• household has at least three members (*hhsize* ≥ 3), • two members form a married (*rel* = 1), cohabiting (*rel* = 2) or same-sex partnership couple (*rel* = 3), • the couple consists of the biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster (*rel* = 12) parents of all others, • other members are biological (*rel* = 13), step (*rel* = 14, *rel* = 15), adoptive (*rel* = 16) or foster (*rel* = 17) siblings to one another, • at least one of the other members (children) is younger than 18 years (*age* < 18)
7	other	• other households
8	inconsistent household relations	• relations in the household are mutually incompatible

Table 30: Individual-level derived variables

age	age in years derived from month and year of birth (*birthm*, *birthy*) and month and year of completion of the CAPI household questionnaire (*hstartm*, *hstarty*). When the age indicated in self-completion questionnaires and diaries (*papiage*, *diaryage*) differed by one year and fell within the instrument's age limits, the *papiage*/*diaryage* value was imputed as *age*
agecat	age categorized in 10-year intervals
primary	person occurred in the sample primarily (member of unsplit household, current member of split household) or secondarily (leaver from split household)
hhmem(*x*)	household membership in Wave (x)
capi(*x*)	whether household member completed the CAPI individual questionnaire in Wave (x)
capistat	indicator of the CAPI individual questionnaire completion in previous years
capi	indicator of the CAPI individual questionnaire completion in at least one previous wave
livesp*)	whether household contains the person's spouse (*rel* = 1) or registered same-sex partner (*rel* = 3)
livewith*)	whether household contains the person's cohabiting partner (*rel* = 2)
livefa*) **)	whether household contains a man (*sex* = 1) who is the person's biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster (*rel* = 12) father
livemo*) **)	whether household contains a woman (*sex* = 2) who is the person's biological (*rel* = 9), step (*rel* = 10), adoptive (*rel* = 11) or foster (*rel* = 12) mother
livebifa*)	whether household contains a man (*sex* = 1) who is the person's biological father (*rel* = 9)
livebimo*)	whether household contains a woman (*sex* = 2) who is the person's biological mother (*rel* = 9)
livegfa*)	whether household contains a man (*sex* = 1) who is the person's grandfather (*rel* = 21)
livegmo*)	whether household contains a woman (*sex* = 2) who is the person's grandmother (*rel* = 21)
bipar*)	whether at least one household member is the person's biological child (*rel* = 4)
child	number of household members who are the person's biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) children
childy	number of household members under 18 years (*age* < 18) who are the person's biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) children
childa	number of adult (*age* ≥ 18) household members who are the person's biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) children
etran	change of economic status between waves
relup	change of residential partner between waves
own	whether household member is among the owners of the property in which the household resides
coo	whether household member is among the members of the cooperative that owns the flat in which the household resides
part	whether CAPI individual questionnaire respondent has a resident (*livesp* = 1 or *livewith* = 1) or non-resident partner (*date* = 1) or is married or in same-sex partnership (*mstat* = 1 or *mstat* = 2). When value could not be determined due to invalid answers to source variables, the value of *papipart* in the adult self-completion was imputed as *part.*
sumchild	total number of biological (*rel* = 4) and adoptive (*rel* = 6) residential or non-residential (*nrechn*) children of CAPI individual and proxy questionnaire respondents. In Wave 3, the *nrechn*

	variable was included in the CAPI individual questionnaire only for those who had never completed this instrument before (*capi1* = 2 and *capi2* = 2). The value of the variable *sumchild* in Wave 3 is for this reason computed from the *nrechn* variable from the year the respondent last completed the CAPI individual questionnaire.
nage(*x*)	age of non-residential children derived from the *nbirthm(*x*)* and *nbirthy(*x*)* variables
mrenum	number of reasons to relocate (*mwork*, *mfam*, *mhous*, *melse*) selected by CAPI individual questionnaire respondents

*) The *live*- and *bipar* variables are derived from relations between household members (the *rel*- variables) and can only equal 1 (yes) or 2 (no). Therefore, these variables do not indicate whether there is someone outside the household who is related to the person in the given way (e.g., the value "no" of the *livesp* variable means that no one in the household is the person's spouse, not that the person does not have a spouse living outside the household).

**) For persons who share the household with a step parent of the same sex as their biological parent, the value "yes" is also set for the variable related to the opposite sex to that of the biological parent.

10. Stratification status

Values of the ISEI scale (International Socio-Economic Index)[6] were assigned to the variables *isco08*, *fisco08*, *misco08, aisco08* and *eisco08*.

The ESeC indicator (European Socio-economic Classification)[7] was derived from the variables *estat (festat, mestat)*, *jobtp*, *svise* (*fsvise*, *msvise*), *empl* (*fempl*, *mempl*), *empls* (*fempls*, *memps*) and *isco08* (*fisco08*, *misco08*).

ISEI values are available for individual CAPI respondents, individuals for whom the proxy CAPI was completed, and child self-completion respondents. ESeC values are available for individual CAPI respondents and individuals for whom the proxy CAPI was completed.

The SPSS syntax used in the construction of the ISEI and ESeC indicators is available as a separate file.

Table 31: Stratification status indicators

ISEIisco08	respondent's socioeconomic index
ISEIfisco08	socioeconomic index of respondent's father (when respondent was 16 years old)
ISEImisco08	socioeconomic index of respondent's mother (when respondent was 16 years old)
ISEIaisco08	socioeconomic index of child's aspirational future job
ISEIeisco08	socioeconomic index of child's expected future job
ESeC08	9-class ESeC, respondent
ESeC08f	9-class ESeC, respondent's father (when respondent was 16 years old)
ESeC08m	9-class ESeC, respondent's mother (when respondent was 16 years old)
ESeC08_6	6-class ESeC, respondent
ESeC08f_6	6-class ESeC, respondent's father (when respondent was 16 years old)
ESeC08m_6	6-class ESeC, respondent's mother (when respondent was 16 years old)

[6] Ganzeboom, Harry B. G.; Treiman, Donald J., *International Stratification and Mobility File: Conversion Tools.* Amsterdam: Department of Social Research Methodology, http://www.harryganzeboom.nl/ismf/index.htm. <January 12, 2010>.

[7] Harrison, Eric. *European Socio-economic Classification.* http://ekharrison.weebly.com/european-socio-economic-classification-esec.html <February 5, 2015>.

11. Individual income variables

Individual income questions from various sources (individual and proxy CAPI) were used to construct several summary variables, namely employment income, business income, income from welfare benefits, income from pensions, and income from capital assets and other sources.

Income in all of the above categories is indicated as monthly amounts. Whereas the figures for employment, pension and social welfare income are actual monthly amounts (monthly amounts indicated by the respondent or annual amounts divided by the number of months of receiving income), the figures for business and capital income represent one-twelfth of the respondent-indicated annual amounts.

When the respondent had no income of a type, the value 0 was imputed for the relevant derived variables. In cases when the variable could not be constructed due to invalid answers to at least one of the source variables, the value 98/999,998 was imputed.

Employment income is indicated both for individual CAPI respondents and for household members for whom a proxy interview was conducted. All other income variables can only be computed for individual CAPI respondents.

Employment income is indicated both in the form of amounts (*dignum* and *dinnum* variables) and as an categorical variable with income ranges (*digcam* and *dincam* variables). Only one type of information was stated by the respondents, whereas the other type was derived. Individual CAPI respondents were preferably asked about exact amounts, and only answered the income interval question if they refused to state or did not know the exact amount. Only the income interval question was asked in the proxy interviews. The exact amounts were derived from the interval variable as central values of the given interval (e.g., CZK 2,500 is the central value of the CZK 1–4,999 interval), except the top intervals, where the median of the values stated by respondents falling within that interval was imputed.

The *c_dignum* and *c_dinnum* variables contain the source information for the *dignum* a *dinnum* variables.

Table 32: Construction of the *dignum* and *dinnum* variables

value	questionnaire	mode of construction
0	CAPI, individual	the *payg/payn* variable indicates 0
1	CAPI, individual	the *payg/payn* variable indicates a monthly amount (*paygam* = 2)
2	CAPI, individual	the *payg/payn* variable indicates an annual amount (*paygam* = 1), monthly amount calculated as annual income divided by number of months of receipt in the *paymon* variable
3	CAPI, individual	don't know or refusal in the *payg/payn* question, but a valid answer to *paygcat/payncat*. Central value of the income interval for the category imputed; median income imputed for the top interval.
4	CAPI, proxy	only the *paygcat/payncat* categorized income question asked. Central value of the income interval for the category imputed; median income imputed for the top interval.

The *signum* and *sinnum* variables indicate gross and net monthly **business income.** In the absence of a question on number of months of receiving business income, the monthly amount was computed as one-twelfth of the annual income. Positive amounts indicate business profit, negative ones indicate a loss.

The *penum* variable indicates **income from pensions.** Included are old-age, widow, widower, orphan and disability pensions. The amount equals the sum of the monthly amounts in source variables.

The total monthly amount of the different **social welfare** variables (unemployment benefits, Assistance in Material Need, sickness benefits, maternity allowance, parental allowance, and stipends) is indicated by the *benum* variable.

The *cognum* variable equals the total **income from capital assets and any other income** not included in the above categories. The monthly amount is calculated as one-twelfth of the annual income.

Table 33: Individual income variables

		questionnaire	source variables
dignum	amount of gross monthly employment income, in CZK	CAPI, individual and proxy	*payg, paygam, paymon, paygcat*
c_dignum	source information for the construction of the *dignum* variable	CAPI, individual and proxy	*payg, paygam, paymon, paygcat*
dinnum	amount of net monthly employment income, in CZK	CAPI, individual and proxy	*payn, paynam, paymon, payncat*
c_dinnum	source information for the construction of the *dinnum* variable	CAPI, individual and proxy	*payn, paynam, paymon, payncat*
digcam	categorized gross monthly employment income, in CZK	CAPI, individual and proxy	*payg, paygam, paymon, paygcat*
dincam	categorized net monthly employment income, in CZK	CAPI, individual and proxy	*payn, paynam, paymon, payncat*
signum	amount of gross monthly business income, in CZK (1/12 of annual amount)	CAPI, individual	*sfpayg, gprof*
sinnum	amount of net monthly business income, in CZK (1/12 of annual amount)	CAPI, individual	*sfpayn, nprof*
penum	amount of monthly income from pensions (old-age, survivor, disability), in CZK	CAPI, individual	*opena, wpena, dpena*
benum	amount of monthly income from welfare benefits (unemployment benefits, Assistance in Material Need, sickness benefits, maternity allowance, parental allowance, stipends), in CZK	CAPI, individual	*ucoma, nebena, sibena, mabena, pallowa, schola*
cognum	amount of gross monthly income from capital assets and other sources, in CZK (1/12 of annual amount)	CAPI, individual	*capina, othina*

12. Interview context variables

The datasets contain a set of variables accounting for the context and procedure of interviewing (so-called paradata), which can be divided into four categories:

- household's place of residence,
- interview characteristics,
- interview date, and
- completion of questionnaires in the household.

In the data, household's place of residence is specified at the level of regions (NUTS 3) and counties (LAU 1). Also included are the *strata* and *psu* variables indicating units used in the sampling process. *Strata* (1 to 58) indicates a combination of the size of municipality of residence (*msize*) and the region of residence (*region*). *Psu* indicates groups of addresses that comprise the dwellings from the first stage of sampling and addresses in their geographic proximity from the second stage (the sampling procedure is described in Section 2.3, *Sampling the households*).

Municipality size (*msize*), *region* and *county* indicate the address at the time of interviewing, whereas *strata* and *psu* refer to the time of sampling (June 2015). However, several households approached in Wave 1 relocated before fieldwork and were interviewed at their new address. These households have a different place of residence than other households in the same primary sampling unit or stratum.

For households that did not relocate, the data from all waves contain the same municipality size according to 2015 data, even if official statistics indicate that the municipality shifted into another size category.

Table 34: Household's place of residence

msize	municipality size
region	region (NUTS 3)
county	county (LAU 1)
strata	stratum (1 to 58), Wave 1 only
psu	primary sampling unit (1 to 1275), Wave 1 only

The data contains indicators for distinguishing between interviewing modes (CAPI = personal interviewing or self-administered questionnaires, CAWI = web-based completion of all instruments), the agency that interviewed the household, and the interviewer. Interviewer number is given only for households that took part in personal interviewing, not CAWI. The *intid* variable corresponds with the interviewer ID from contact forms and interviewer datasets.

Also included are a few indicators of type of instrument completed, interviewer-declared reason for completing a proxy interview, and derived relation between the proxy respondent and the person he/she accounts for.

Table 35: Interview characteristics

qmode	whether household took part in personal or CAWI interviewing
agen	agency that interviewed the household
intid	interviewer ID (the same interviewer has the same ID across all survey waves)
iorder	order of completion of the CAPI individual questionnaire in the household (only for respondents with whom the CAPI questionnaire was completed in person, not by proxy)
proxy	whether the CAPI interview only took the reduced proxy form
relpro	relation of proxy respondent to the person for whom the questionnaire was completed
reaspro	reason for conducting a proxy interview
papi	type of self-administered questionnaire completed (only for respondents who completed a self-administered questionnaire)

Days and months of starting and ending the CAPI interview were recorded automatically by the CAPI software. The respondents entered this information themselves in the self-administered questionnaires.

Table 36: Interview date

hstartm	month of starting the CAPI household questionnaire
hstartd	day of starting the CAPI household questionnaire
hendm	month of ending the CAPI household questionnaire
hendd	day of ending the CAPI household questionnaire
istartm	month of starting the CAPI individual questionnaire
istartd	day of starting the CAPI individual questionnaire
iendm	month of ending the CAPI individual questionnaire
iendd	day of ending the CAPI individual questionnaire
pstartmo	month of starting the CAPI proxy questionnaire
pstartd	day of starting the CAPI proxy questionnaire
pendm	month of ending the CAPI proxy questionnaire
pendd	day of ending the CAPI proxy questionnaire

papimon	month of completing the self-administered questionnaire
papiday	day of completing the self-administered questionnaire

The dataset indicates the numbers of instruments completed in each household. The shares of instruments completed represent the ratio of the number of instruments completed to the number of household members of defined age. The value of child instruments-related variables is system-missing for households in which there were no children aged 10 to 17 years.

Table 37: Completion of questionnaires in the household

h_indi	number of CAPI individual questionnaires completed in the household
hp_indi	share of CAPI individual questionnaires completed in the household
h_indi_proxy	number of CAPI individual and proxy questionnaires completed in the household
hp_indi_proxy	share of CAPI individual and proxy questionnaires completed in the household
h_papi_adult	number of self-administered questionnaires completed for adults in the household
hp_papi_adult	share of self-administered questionnaires completed for adults in the household
h_papi_child	number of self-administered questionnaires completed for children in the household
hp_papi_child	share of self-administered questionnaires completed for children in the household
h_diary_adult	number of time-use diaries completed for adults in the household
hp_diary_adult	share of time-use diaries completed for adults in the household
h_diary_child	number of time-use diaries completed for children in the household
hp_diary_child	share of time-use diaries completed for children in the household

13. Variables routed based on another person's values

The CAPI individual questionnaires contain several questions that are routed on the basis of another household member's characteristics, i.e. the routing rule is derived from variable values pertaining to another person. Such questions were administered to parents when their cohabiting child met defined conditions, and the answers referred to that child.

When a routing rule refers to the value of a variable for a household member whose *pno* = *x* then the prefix (x) is used in that variable's code. This notation should be distinguished from the suffix (x), which indicates a looped variable of order *x*.

For example, the routing rule applied to the *imtalk2* variable, "*rel2* = 4 | 5 | 6 | 7 AND *2age* >= 10 AND *2age* <= 17", means that the question will be asked if the respondent is the parent of the 2nd person in the household and that person is aged 10–17 years.

Table 38: Variables filtered based on another person's values

imtalk(*x*), quar(*x*), hapschw(*x*)–hapliv(*x*)	asked of biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) parents whose child is aged 10 to 17 years (*age* of child ≥ 10 and ≤ 17). Number in variable code indicates order of child in household (*pno*).
chtime(*x*)	asked of biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) parents of child. Number in variable code indicates order of child in household (*pno*).
pdisch(*x*)	asked of biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) parents whose child is at least 8 years old (*age* of child ≥ 8). Number in variable code indicates order of child in household (*pno*).
eduas(*x*)	asked of biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) parents whose child is younger than 25 years (*age* of child < 25) and studying (*std* of child = 1). Number in variable code indicates order of child in household (*pno*).
eduex(*x*)	asked of biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) parents whose child is younger than 25 years (*age* of child < 25) and studying (*std* of child = 1). Number in variable code indicates order of child in household (*pno*).
raidh(*x*)	asked of biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) parents whose child is at least 15 years old (age of child ≥ 15) and owns the household's housing property (*own* = 1) or is a member of the cooperative (*coo* = 1). Number in variable code indicates order of child in household (*pno*).
rplaidh(*x*)	asked of biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) parents whose child is at least 15 years old (age of child ≥ 15) where, if he/she owns the household's housing property (*own* = 1) or is a member of the cooperative (*coo* = 1), the parent has not yet provided it with financial assistance to obtain housing (*raidh* = 4 | 88 | 99), or he/she does not own the current housing property or is a member of the cooperative (*raidh* = 97). This routing condition applies to Wave 2, it was changed for Wave 4. Number in variable code indicates order of child in household (*pno*).
maben	asked of biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) parents who have in the household at least one child aged up to 2 years (*age* of child ≤ 2)

pallow	asked of biological (*rel* = 4), step (*rel* = 5), adoptive (*rel* = 6) or foster (*rel* = 7) parents who have in the household at least one child aged up to 4 years (*age* of child ≤ 4)

14. Data anonymization

The instruments did not contain questions about respondents' personal data. Personal data (name and surname, telephone and e-mail contacts) were only collected for the purposes of repeated contact and including the households in a lottery. Household members provided that information using a separate form, where they gave their written consent to use of the data for the defined purposes only. Since personal data could be linked to questionnaire data, Wave 4 fieldwork (2018) included obtaining households' written consent to interviewing and parental written consent to their children's interviewing.

Selected variables that provide detailed information about the respondent were not included in the data (for an overview, see Section 5.3, *Variables not included in the dataset*).

Numerical household and person IDs (*hid*, *pid*) from primary data were replaced with new random IDs in accordance with the identifier construction rules described in Section 6, *Identification variables*.

The values of variables *nname1* to *nname8* (given names of non-residential children) were replaced by numerical codes linked to a list of first names.

15. Data checks, processing and cleaning

15.1. Data checks

Several mechanisms to enhance validity and consistency of the data collected were built into the CAPI instruments. For several questions with numerical answers in the CAPI household and individual questionnaires, checks were implemented that prompted the interviewer to verify the numerical value entered if it did not fall within a defined interval (see questionnaires for details). In variables identifying the relations between household members (*rel*-), entry of one person's relation to another was followed by automatic imputation of the latter's relation to the former. In the introductory part of the individual CAPI, eight basic sociodemographics were verified, and any inaccuracies could be corrected. The goal was to check whether the individual CAPI was being completed with the right respondent, and to make sure the variables key to subsequent routing were correct. In the CAPI questionnaires and the CAWI mutations of all instruments, no question could be left unanswered.

In the field, interviewers verified the formal completion accuracy of self-administered questionnaires and diaries, asking the respondents to correct any mistakes. They also made sure that a large number of questions (or timeslots) were not left answered in the self-administered questionnaires and diaries.

With weekly (Wave 1) or biweekly (Wave 2) periodicity, the data was exported from the interviewing software to SPSS, and the fieldwork agencies along with the research team participated in ongoing checks of consistency between data and the data collection instruments, including, in particular, code definitions for individual responses, agreement with a defined range of values, and routing rules. Numerical variables without a defined range of permissible values were checked for extreme values. For any extreme values found, it was verified whether they had been entered by interviewers or resulted from data processing errors. Variables in the form of certain types of events (e.g., *madres*) were checked for response consistency. Although mutually inconsistent years or improbable numerical values were identified in the process of checks, there was no respondent for whom such inconsistencies would accumulate to an extent suggesting his/her answers as a whole were systematically incongruent or outright random. In these cases, the values were left unchanged.

From Wave 2, the checks also focused on cross-wave consistency of the different household members' sociodemographics. In case of a major inconsistency (typically when more than two stable characteristics changed), the interviewer was consulted about the household (this could only be done for CAPI data) and the household excluded from further processing unless the changes were explained satisfactorily. During fieldwork monitoring, the cross-agency consistency of distributions was checked. If any differences between the two fieldwork agencies were found, possible causes were examined, and harmonized interviewing procedures were put in place. Both the data batches coming in during fieldwork and the final dataset were checked for consistency with the instruments, extreme values and consistency of distributions.

In the data entry process for self-administered questionnaires (manual entry by STEM/MARK, scanning by MEDIAN) and diaries (manual entry by both agencies), the instruments were checked for the level of completion and linked to CAPI data using respondent IDs. STEM/MARK performed double data entry to verify 20% of the self-completion questionnaires and diaries data in all waves. MEDIAN checked the

scanning settings for self-completion questionnaires and performed manual data checks for several dozen randomly drawn questionnaires. In addition, it checked all items for which the scanning yielded no answer, with manual entry of any valid values. MEDIAN did not use double data entry to verify the diaries in Wave 1 but did apply the procedure to 100% of the diaries from Wave 2.

15.2. Excluded questionnaires

The CAPI individual and proxy questionnaires, self-completions and diaries were checked for compliance with respondent age limits. Self-administered instruments completed by age-ineligible respondents were excluded. If the values of age reported in the self-completion (*papiage*) and the household CAPI (*age*) differed by 1 year at most and, at the same time, the self-completed age was consistent with the instrument's age limit, then the value of *papiage* was imputed as *age.*

In Waves 2 to 4, a few self-completion questionnaires and diaries whose identification variables referred to leavers were excluded.

Self-completion questionnaires and diaries that could not be linked to other instruments due to missing IDs were excluded from processing. This was the case for tens of cases per wave or fewer.

Parental consent with child instrument completion was sought in Wave 4. A small number of self-completions and time-use diaries were excluded from data processing due to missing consent.

15.3. Dataset production

The fieldwork agencies provided the research team with separate datasets for each instrument (in addition, in Wave 1, the household CAPI was divided into two datasets with household-level and individual-level variables, respectively). The different datasets were then merged into a single file. When the same question was answered in more than one instrument (e.g., the individual CAPI and the child self-completion), only one variable was included in the dataset (for more on the structure of the dataset, see Section 5, *Dataset content and structure*).

15.4. Data cleaning

Relations within the household (the *rel1* to *rel12* variables) were checked for consistency of answers provided by the different household members. Mutually inconsistent or otherwise incongruent relations (e.g., a father younger than his son) were verified and corrected in line with interview recordings or with the logic of other relations and demographic variables.

The variables *name*, *sex*, *birthm*, *birthy* and, in part, also *rel1* to *rel12* (permanent relations) were cleaned in line with the values corrected by respondents in subsequent waves (from Wave 2, the household CAPI prompted the respondents to correct information entered in the previous waves), and variables routed based on them were modified accordingly.

The variables *mstat*, *estat*, *educ*, *std*, *stdch* and *rel1* to *rel12o2* were cleaned individually in line with the values of contingent variables in the same wave (e.g., in response to the open-ended question *stype,* the respondent stated that he/she was in fact not studying), the values of the same variable entered in the following waves, and subsequent waves' values of contingent variables (e.g., *mstat* was cleaned in line with *mchange* values), considering the overall consistency of the different answers. Changes made to underlying sociodemographics were followed by the cleaning of variables routed on their basis.

For the *ustarty(*x*)*, *uendy(*x*)*, *pstarty* and *lstarty* variables, invalid answers (don't knows and refusals) were recoded as 999,996 if the last wave's *estat* value was equal to the current *estat*. For these cases, the answers to the questions *ustartm*, *uendm*, *pstartm* and *lstartm* were corrected to 96.

For the *mstat*, *birthy* and *age* variables, highly unlikely answers (born before 1900, marital status same-sex partnership when other indications suggested cohabitation) were recoded as 98/999,998.

For the *svise* and *empl* variables, the value 1 was recoded to 2 if the respondent gave a zero answer to *subor* or *svise*, respectively.

Values over 140 were replaced with an error code 999,998 in variables *shours, hours, adhours, choosehr* and *hswork.* This code was also used to replace values under 100 and over 250 in the *height* variable, and values under 40 and over 250 in the *weight* variable.

If the routing variables *papipart* and *papijob* were left unanswered but the respondent gave valid answers in the following questions, *papipart* was recoded to the value 1 if *part* = 1, and *papijob* was recoded to 1 if *estat* = 1/2/3/4/5/6/9 or *job* = 1.

The values of *asame* variable were compared to household addresses[8] from all previous waves, and any inconsistencies (e.g., the same address in both waves when *asame* indicated a change) were reconciled through consultation with the interviewers. Interviewers were also consulted on all changes of address suggested by the *asame* value to verify if that change referred to actual relocation or merely a correction to the address previously entered. In the latter case, the *asame* variable was recoded as 1 (same address) and the contingent variables were cleaned.

Variables concerning the day and sleep regime from the adult self-completion in Wave 4 (*dayreg* to *daylifm*) were cleaned in order to satisfy the formal principles of their completion, i.e. so that invalid codes 96, 98 and 99 apply both to the hour and minute parts of the answer. Answers falling outside the valid ranges defined by the possible number of days, hours and minutes and answers not meeting the required completion format were cleaned in the manner shown in *Table 39* (the last column contains the number of cases with at least one instance of cleaning of the given type performed). In variables representing the number of hours, values 24 were kept only in the variables *dayliwh* and *daylifh* which express duration, not time point. In other variables referring to the number of hours, values 24 were recoded to 0.

[8] Household addresses are kept in the interviewing software, separately from the CAPI questionnaires. From Wave 2, interviewers were able to enter a new household address in the system.

Table 39: Cleaning of variables dayreg to daylifm

variables	initial values before cleaning	values after cleaning	no. of cases
dayreg	> 7	98	27
bedwh, sleepwh, bedfh, sleepfh	H = 5, 6, 7, 8, 9, 10, 11, 12 BS > W or W – BS < 4 *)	H = 17, 18, 19, 20, 21, 22, 23, 0	80
bedwh, sleepwh, wakewh, bedfh, sleepfh, wakefh	H = 24 **)	H = 0	840
bedwh, sleepwh, wakewh, bedfh, sleepfh, wakefh	H > 24	H = 98	9
bedwm, sleepwm, wakewm, bedfm, sleepfm, wakefm	M = 99 H < 24	M = 0	153
bedwm, sleepwm, wakewm, bedfm, sleepfm, wakefm	M > 60	M = 98, H = 98	1
bedwm, sleepwm, wakewm, bedfm, sleepfm, wakefm	M = 60	recalculated to hours	0
mactish, mactieh	H = 24	H = 0	34
mactish, mactieh	H > 24	H = 98, M = 98	2
mactism, mactiem	M = 99, H < 24	M = 0	114
mactism, mactiem	M > 60	M = 98, H = 98	0
mactism, mactiem	M = 60	recalculated to hours	1
dayliwh, daylifh	H > 24	H = 98, M = 98	65
dayliwm, daylifm	M = 99, H <= 24	M = 0	119
dayliwm, daylifm	M >= 60	recalculated to hours	15
dayliwh, daylifh	H = 99	H = 0	6

*) BS = bedwh, sleepwh, bedfh, sleepfh, W = wakewh, wakewh, wakefh, wakefh

**) H = answer part representing hours, M = answer part representing minutes

15.5. Coding open-ended and semi-closed questions

In the open-ended and semi-closed questions *newre, leftre, rel, lang1, lang2, tenelse, ownelse, cooelse, dwelob, opet(x), nsend, stype, send, stype2, edusec, eduprim, olngtp, jobnot, jendre, jwhy, jlwhat(x), unemre, partycl, partyvt7, partywv, helpty(x), ohelpty, rework, refam, rehous, reelse, mdwelob, hendre(x), unre(x), cborn, rgborn, citot, nation, relig, ashigh, exelse* and *cpartywv*, the open-ended answers were coded into broader categories under a string variable or assigned to corresponding closed-ended options. The answer options of the new string variable recoded from the open-ended accounts are not necessarily mutually exclusive or exhaustive, since they largely respect the content of and level of detail provided in the original accounts.

In relevant cases, open-ended accounts of the reason for relocation from a former place of residence (rework, refam, rehous, reelse) were assigned to another existing category of relocation reasons.

Job description variables (swrkps, nwrkps, wrkps, pwrkps, hwrkps(x), fwrkps, mwrkps, jobas, jobex) were coded in line with the CZ-ISCO classification.

15.6. Value imputation

In the individual CAPI variables *htnr*, *isco08*, *svise*, *subor*, *empl*, *empls*, *fsize* and *contr* of Wave 2 to Wave 4, Wave 1 values were imputed if the respondent confirmed their validity in dependent interviewing (see Section 8, *Dependent interviewing variables*).

No data imputation was performed for the self-completion questionnaires.

For the diaries, respective timeslots were coded as "sleep" (1) if the respondent's account ended in evening hours and no longer continued, or if the account only began in the second or higher-order timeslots of the next morning. Imputation is indicated in the diary data. Details are provided in a separate diary data documentation file.

15.7. Adjustments to number of cases based on subsequent waves

After Wave 2 fieldwork was completed, nine individuals were added to Wave 1 data who had already lived in the households surveyed during Wave 1 but had not been accounted for by the interviewer. At the same time, 41 households were excluded from Wave 1 data for which major errors in the interviewing process were found during Wave 2 (shown as "interviewed, excluded by quality control" in *Table 3*, *Field outcomes for the household sample, Wave 1*).

Following Wave 3 fieldwork, 29 additional unaccounted-for individuals were added to Wave 1 data and 15 to Wave 2 data. No households were excluded due to interviewing errors after Wave 3.

Based on Wave 4 fieldwork, 6 cases were added to the datafile from the first wave, 7 to the data from the second wave and 8 to the file form the third wave.

15.8. Invalid answers/values

All questions in the CAPI questionnaires (except the cognitive tests included in Wave 2 CAPI) allowed the interviewer to account for a situation when the respondent was unable or refused to answer the question. In the face-to-face interview, these options were not prompted by the interviewer or included in showcards; however, they were shown after the valid options in the CAWI mutation. In the published CAPI questionnaires, the unprompted options are shown in square brackets.

Some questions in the self-administered questionnaires and their CAWI mutations explicitly allowed the "don't know" option. However, for selected questions (e.g., the political knowledge items *knowkal* to *knowdsss*), the "don't know" option is not coded as invalid answer (88).

During data entry and cleaning, error codes 98/999,998 were imputed for CAPI and self-completion variables in cases referred to in *Table 40*.

In the SPSS data files, invalid codes are defined as user-missing. The user-missing property was also assigned to codes 96/999,996 for the variables *sentryy, sentrym, ustartm*(x)*, ustarty*(x)*, uendm*(x)*, uendy*(x)*, pstartm, pstarty, dcmoney* to *dcsave, cless(x), relparnl, fowny, fcoopy, hendy*(x)*, mleavee*(x)*, uneme*(x)*, nsentryy*(x)*, nsetrym*(x)*, bedwh, bedwm, sleepwh, sleepwm, wakewh, wakewm, bedfh, bedfm, sleepfh, sleepfm, wakefh, wakefm, symods* to *symsso*. Although they were valid prompted options, they refer to answers outside the main range of options. The user-missing property was not assigned to "don't know" answers to self-completion variables unless they were coded as invalid (88).

Table 40: Invalid answers and their coding in the CAPI and self-completion data

88 / 888,888	CAPI	don't know, can't say – unprompted *Included among the options in the CAWI questionnaire.*
	self-completion	don't know (coded as invalid only for selected variables)
99 / 999,999	CAPI	refused – unprompted *Referred to in the CAWI questionnaire as „Don't want to say".*
	self-completion	no answer
97 / 999,997	CAPI/self-completion	the respondent was routed away
98 / 999,998	CAPI	questions not asked due to ill-defined routing rules (which were corrected during cleaning)
	CAPI	*ohousch, nown, part*, *sumchild, dignum, c_dignum, dinnum, c_dinnum, digcam, dincam, signum, sinnum, penum, benum, cognum* = derived variable cannot be computed due to invalid answers to some of the underlying variables
	CAPI	*birthy*, *age*, *mstat* = wrong answer (born before 1900, marital status same-sex partnership when other indications suggested cohabitation)
	CAPI/self-completion	*shours, hours, adhours, choosehr, weight, height, hswork* = answer given was outside the valid range
	CAPI/self-completion	*isco08*, *fisco08*, *misco08*, *aisco08*, *eisco08* (including derived ISEI and ESeC indicators) = cannot code underlying open-ended answer
	self-completion	ambiguous answer (e.g., two answers selected without marking which one applies; illegible answer to open-ended question)

In Waves 1 and 2, the interviewing scripts for the household and individual CAPI were affected by several technical problems. As a result, some questions were not asked, and some answers were not saved, for all relevant respondents. These problems were corrected during Wave 2 fieldwork, and the missing values were established through telephone interviewing in September and October 2016.

The missing values arising in this way are coded as 98/999,998 in the data.

Table 41: CAPI variables with missing values due to script errors		W1	W2
estat	on one type of interviewer laptop computers, the categories 8, 9, 10 and 11 were only visible in landscape view; as a result, frequencies of the visible categories of 12 and 14 were overreported. Changes to economic status from categories 8, 9 and 11 to 12 and 14 were reversed to 8, 9 and 11, respectively.	-	1
exutil, exrent, exmort, exrep	*ohousch* filter variable miscalculated	-	1
chben	*p0to26* filter variable ill-defined for some relevant households	1	-
emplch	routing rule implementation error (ignored condition *w1_empl* = 2)	-	1
svise, subor, fsize	some values not saved – error in communication between questionnaire and FTP for *agen = 0*	1	-
jbsat	routing rule implementation error (ignored condition *estat* = 4)	-	1
inval	routing rule implementation error (ignored conditions *inot* = 1 and *inhous* = 1)	-	1
rplaidh(*x*)	routing rule implementation error (ignored condition *raidh* = 97)	-	1
nplaidh(*x*)	routing rule implementation error (ignored condition *naidh* = 97)	-	1

16. Sample specifics

Compared to 2011 census data[9], one-person households were underrepresented in the CHPS data (22% vs 33%). This may be due to lower chances of contacting those households (lower odds of the single member's being at home compared to the odds of any adult being at home in multiple-member households), their lower willingness to participate (e.g., elderly individuals living alone were afraid of the interviewer), the interviewer incentive system (rewarding the number of instruments completed in a household), or under-splitting of dwellings into households on common budget.

The complete sample of household members can be divided into subsamples depending on completion of the different types of instruments (e.g., the subsample of individual CAPI respondents). These subsamples exhibit structural differences both from one another and especially from the complete sample. Deviations are due to the fact that respondents were not selected randomly but with a view to interviewing all household members. Combined with the fact that within-household response rates were not 100%, this results in overrepresentation of better-accessible sociodemographic segments of the population (women, elderly people) compared to the sample of all household members.

[9] Population Census 2011, Czech Statistical Office. Results available at: https://www.czso.cz/csu/czso/population-censuses

17. Post-stratification weights

17.1. Binary variables indicating instrument completion

The dataset contains post-stratification weights to correct deviations from population proportions in terms of sex, age, education and region of residence. Moreover, weights to achieve a uniform distribution of days in a week are available for time-use diary respondents.

Since the different instruments and their combinations were completed by different groups of respondents, 16 variables with weights related to the different instruments and their combinations were prepared. For the 6 subsamples of respondents who completed the time-use diary, the weights were computed in two versions (the first version only weights on the basis of sociodemographic variables, the second version also ensures uniform distribution of weekdays). In total, there are 22 weighting variables in the data.

Selection of the different respondent groups relies on the binary variables *hh_hh* to *child_papi_diary*, which indicate whether a household member completed the instrument (no matter if he/she fell within the instrument's target population). If yes (1), the respondent obtains a value of the weighting variable for the given instrument or combination of instruments. *Table 42* contains an overview of binary indicators and related weights, *Table 45* illustrates the construction of respondent subsamples.

For leavers (*rstat* = 5) in Wave 2 dataset, these binary variables are coded as 0 and the weighting variables are not computed.

Table 42: Indicators of respondent groups and corresponding weights

hh_hh	household member who completed the CAPI household questionnaire
W_hh_hh	weight for households
hh_indi	member of the surveyed household (i.e. all members except leavers)
W_hh_indi	weight for household members
indi	household member who completed the CAPI individual questionnaire
W_indi	weight for CAPI individual questionnaire respondents
indi_proxy	household member who completed the CAPI individual questionnaire OR for whom a CAPI proxy questionnaire was completed
W_indi_proxy	weight for CAPI individual questionnaire respondents OR respondents for whom a CAPI proxy questionnaire was completed
papi_adult	household member who completed the adult self-completion
W_papi_adult	weight for respondents of the adult self-completion
papi_child	household member who completed the child self-completion
W_papi_child	weight for child self-completion respondents
adult_child	household member who completed the adult self-completion OR the child self-completion
W_adult_child	weight for adult or child self-completion respondents

adult_child15	household member who completed the adult self-completion OR the self-completion for children aged 15–17
W_adult_child15	weight for respondents of the adult self-completion OR the self-completion for children aged 15–17
indi_child15	household member who completed the CAPI individual questionnaire OR the self-completion for children aged 15–17
W_indi_child15	weight for respondents of the CAPI individual questionnaire OR the self-completion for children aged 15–17
diary_adult	household member who completed the adult time-use diary
W_diary_adult	weight for adult time-use diary respondents
Wd_diary_adult	weight for adult time-use diary respondents with uniform distribution of days
diary_child	household member who completed the child time-use diary
W_diary_child	weight for child time-use diary respondents
Wd_diary_child	weight for child time-use diary respondents taking into account the distribution of weekdays
indi_papi	household member who completed the CAPI individual questionnaire AND the adult self-completion
W_indi_papi	weight for respondents of the CAPI individual questionnaire AND the adult self-completion
indi_diary	household member who completed the CAPI individual questionnaire AND the adult time-use diary
W_indi_diary	weight for respondents of the CAPI individual questionnaire AND the adult time-use diary
Wd_indi_diary	weight for respondents of the CAPI individual questionnaire AND the adult time-use diary with uniform distribution of days
adult_papi_diary	household member who completed the adult self-completion AND the adult time-use diary
W_adult_papi_diary	weight for respondents of the adult self-completion AND the adult time-use diary
Wd_adult_papi_diary	weight for respondents of the adult self-completion AND the adult time-use diary with uniform distribution of days
indi_papi_diary	household member who completed the CAPI individual questionnaire AND the adult self-completion AND the adult time-use diary
W_indi_papi_diary	weight for respondents of the CAPI individual questionnaire AND the adult self-completion AND the adult time-use diary
Wd_indi_papi_diary	weight for respondents of the CAPI individual questionnaire AND the adult self-completion AND the adult time-use diary with uniform distribution of days
child_papi_diary	household member who completed the child self-completion AND the child time-use diary
W_child_papi_diary	weight for child self-completion AND time-use diary respondents
Wd_child_papi_diary	weight for child self-completion AND time-use diary respondents with uniform distribution of days

17.2. Weighting method

Weighting was performed using the *rake* function in R (*survey* package), an iterative procedure to align the sample's weighted marginal distributions with known population distributions. Maximum weight value was set as 4 (the *trim* function). Trimming was applied to the weights *Wd_diary_adult*, *Wd_indi_diary*, *Wd_adult_papi_diary*, *Wd_indi_papi_diary* in Wave 1 dataset; *W_indi*, *W_indi_child15*, *Wd_diary_adult*, *W_indi_papi*, *W_indi_diary*, *Wd_indi_diary*, *Wd_adult_papi_diary*, *W_indi_papi_diary*, *Wd_indi_papi_diary* in Wave 2 dataset; *W_papi_adult, W_adult_child, W_adult_child15, Wd_diary_adult, W_indi_papi, W_indi_diary, Wd_indi_diary, Wd_adult_papi_diary, Wd_indi_papi_diary* in Wave 3 dataset; and *W_indi, W_indi_proxy, W_papi_adult, W_adult_child, W_adult_child15, W_indi_child15, W_diary_adult, Wd_diary_adult, W_indi_papi, W_indi_diary, Wd_indi_diary, W_adult_papi_diary, Wd_adult_papi_diary, W_indi_papi_diary, Wd_indi_papi_diary* in Wave 4 dataset.

The R script for the weighting procedure is available as a separate file.

17.3. Variables for weighting

Depending on the instrument, three groups of variables were used for the weighting procedure:

- region of residence (NUTS 3),
- cross-classification of sex, education and age + region of residence (NUTS 3), and
- sex + region of residence (NUTS 3).

The control distributions for weighting were sourced from the Labour Force Survey of the given year (2015 for Wave 1, 2016 for Wave 2 etc.), the Population Census of 2011, and Czech Statistical Office data on age distribution of the population on 1 July of the given year.

Table 43: Groups of variables for weighting

1	region of residence (NUTS 3)	**region:** 2011 Population Census (number of households on common budget by region)
2	cross-classification of sex, education and age + region of residence (NUTS 3)	**region:** Age distribution of the population on 1 July of the year (Czech Statistical Office) **sex, education, age:** Labour Force Survey of the year
3	sex + region of residence (NUTS 3)	**region, sex:** Age distribution of the male population on 1 July of the year and Age distribution of the female population on 1 July of the year (Czech Statistical Office)

The variables *sex* and *region* were not modified for weighting. The continuous variable *age* was divided into five or four age categories depending on the instrument's population (resulting in three different age categorizations). The *educ* variable for education was divided into four categories (*Table 44*). The control data from which the theoretical distributions were derived are available as a separate MS Excel file.

The weights for diary respondents were calculated in two variants. The first variant works with the above sociodemographic variables only, whereas the second variant weights not only those but also ensures a uniform distribution of weekdays. The uniform distribution of weekdays was included as a separate criterion, not in cross-classification with the other variables.

Table 44: Categories of variables for weighting

region *(region)*	**1–14**	in line with values of the *region* variable						
sex *(sex)*	**1**	male (*sex* = 1)						
	2	female (*sex* = 2)						
age1 *(ageR)*	**0**	missing value	**age2** *(ageR2)*	**0**	missing value	**age3** *(ageR3)*	**0**	missing value
	1	*age* ≤ 17		**1**	10 ≤ *age* ≤ 34		**1**	15 ≤ *age* ≤ 34
	2	18 ≤ *age* ≤ 34		**2**	35 ≤ *age* ≤ 54		**2**	35 ≤ *age* ≤ 54
	3	35 ≤ *age* ≤ 54		**3**	*age* ≥ 55		**3**	*age* ≥ 55
	4	*age* ≥ 55						
education *(educR)*	**0**	missing value						
	1	lower education, i.e. no education, first or second stage of basic (*educ* = 0, *educ* = 1, *educ* = 2)						
	2	medium education, i.e. secondary vocational (*educ* = 3), lower secondary vocational (*educ* = 4), secondary education with *maturita* exam (*educ* = 5, *educ* = 6, *educ* = 7)						
	3	higher education, i.e. tertiary vocational (*educ* = 8) and tertiary education (*educ* = 9, *educ* = 10, *educ* = 11)						

17.4. Dealing with missing values

Both CHPS and Labour Force Survey data contain missing values in the age and education variables. Missing values were processed in line with the following rules:

1 Missing values in both the control data and the sample data: Cells with missing values are used in standard ways (in none of the cases did the number of missing answers from the sample exceed the number of missing answers in the control data).

2 Missing values in sample data only: The values in the cells with missing answers are copied into corresponding cells of the control table (assuming MAR). The remaining values in the control table are adjusted proportionally so that the total number of cases and the ratio between values in the table's cells remain unchanged.

3 Missing values in control data only: Missing values in the control table are neglected (assuming MCAR). The share of missing values in the control data did not exceed 0.01%.

Table 45: Respondent subsamples and weighting thereof

	Hous. CAPI	Indi. CAPI	Prox. CAPI	Adult self-completion	Child self-completion, age 15–17	Child self-completion, age 10–14	Adult diary	Child diary	n (Wave 1)	n (Wave 2)	n (Wave 3)	n (Wave 4)	Weighting variables	Age cat.
hh_hh	x								5,159	4,147	3,616	3,188	1	-
hh_indi	x								13,083	10,498	9,183	8,093	2	1
indi	&	&							7,118	5,270	4,635	4,021	2	1
indi_proxy	&	+	+						7,605	5,603	4,870	4,222	2	1
papi_adult	&			&					8,131	6,561	5,839	5,132	2	1
papi_child	&				+	+			8,66	602	644	580	3	-
adult_child	&			+	+	+			8,997	7,163	6,483	5,704	2	2
adult_child15	&			+	+				8,437	6,770	6,072	5,344	2	3
indi_child15	&	+			+				7,424	5,479	4 ,868	4,233	2	3
diary_adult	&						&		7,955	6,723	5,966	5,391	2	1
diary_child	&							&	804	602	625	562	3	-
indi_papi	&	&		&					6,339	4,771	4,268	3,603	2	1
indi_diary	&	&					&		6,239	4,819	4,268	3,670	2	1
adult_papi_diary	&			&			&		7,589	6,246	5,627	4,959	2	1
indi_papi_diary	&	&		&			&		5,989	4,603	4,153	3,515	2	1
child_papi_diary	&				+	+		&	778	534	602	531	3	-

x cases when the instrument indicated was completed

\+ cases when at least one the instruments indicated was completed

& cases when all of the instruments indicated were completed

18. Dataset versions

By 18 February 2019, the dataset from CHPS Wave 1 had been published in five versions.

Table 46: Wave 1 dataset versions

	publication date	changes made
v1.0	6 Feb 2017	-
v1.1	6 Mar 2017	excluded the variables *read*, *parnec*, *gchrep*, *gchclean*, *gchmeal*, *gchlaund*
v2.0	11 Apr 2017	• reworded the labels of selected variables in line with those used in Wave 2 dataset • *jobtp* (*agen* = 0): corrected cleaning error (transformed the values 99 to 1/2/88) • *svise*: corrected cleaning error (transformed the values 97 to 1/2/88 in the context of correcting the variable *jobtp*); added the error code 98 • *subor* (*agen* = 0): corrected cleaning error (transformed the values 999,997 to 1–1,000/888,888 in the context of correcting the variable *jobtp*); added the error codes 999,998 • *empl* (*agen* = 0): corrected cleaning error (transformed the values 97 to 1/2 in the context of correcting the variable *jobtp*) • *empls* (*agen* = 0): corrected cleaning error (transformed the values 999,997 to 1–15 in the context of correcting the variable *jobtp*) • *fsize* (*agen* = 0): corrected cleaning error (transformed the values 97 to 1–6/88 in the context of correcting the variable *jobtp*); added the error codes 98 • *fsubor* (*agen* = 0): corrected cleaning error (transformed the values 99 to 1/2/88) • *ESeC08*: changed the value in the context of correcting the variable *jobtp* • *ESeC08_6*: changed the value in the context of correcting the variable *jobtp* • *read* (*agen* = 0, *papi* = 3): corrected errors due to ill-defined scanning procedure • *parnec* (*agen* = 0, *papi* = 2): corrected errors due to ill-defined scanning procedure • *gchrep* (*agen* = 0, *papi* = 3): corrected errors due to ill-defined scanning procedure • *gchclean* (*agen* = 0, *papi* = 3): corrected errors due to ill-defined scanning procedure • *gchmeal* (*agen* = 0, *papi* = 3): corrected errors due to ill-defined scanning procedure

		• *gchlaund* (*agen* = 0, *papi* = 3): corrected errors due to ill-defined scanning procedure • self-completion variables (*agen* = 0): added values that had not been read by scanning (transformed the values 99 to valid ones) • *exelseo*, *cpartywv*: adjusted coding scheme in line with that applied to Wave 2 dataset
v3.0	6 Feb 2018	• renamed the indicators *hid* and *pid* to *w1_hid* and *w1_pid* • adjusted sociodemographic variables in line with Wave 3 data • added 29 household members who had been omitted in Wave 1 • reworded the labels of selected variables • adjusted width of selected string variables in line with Wave 2 dataset • modified calculation of the *part* variable (included adjustment in line with the value of *papipart*) • corrected error in the calculation of the *sumchild* variable • renamed the *pstartm* variable (month of starting the CAPI interviewing) to *pstarmo* because the original name was the same as that of a new Wave 2 variable • added the *intid* variable to indicate interviewer ID • added the *dignum – cognum* variables • added the variables *Wd_diary_adult, Wd_diary_child, Wd_indi_diary, Wd_adult_papi_diary, Wd_indi_papi_diary, Wd_child_papi_diary*
v4.0	18 Feb 2019	• adjusted sociodemographic variables in line with Wave 4 data • added 6 household members who had been omitted in Wave 1 • recalculated weights following the inclusion of omitted household members and adjustments to sociodemographic variables • added value labels not yet included in the datafile and applicable to any of the other waves • harmonized variable formats between waves • reworded the labels of selected variables and values • adjusted the user-missing value setting (changed from discrete values to ranges) • corrected data cleaning based on routing rules (especially *maben*, *maliv*, *faliv*) • corrected minor issues in the coding of string variables (e.g. *cpartywv*) • added variables *nage1* to *nage8* holding ages of non-resident children • corrected calculation of *dinnum* • replaced zeros with system missing values for households with no children aged 10 to 17 years in *h_papi_child*, *h_diary_child*

By 18 February 2019, the dataset from CHPS Wave 2 had been published in two versions.

Table 47: Wave 2 dataset versions

	publication date	changes made
v1.0	6 Feb 2018	-
v2.0	18 Feb 2019	• adjusted sociodemographic variables in line with Wave 4 data • added 7 household members who had been omitted in Wave 2 • recalculated weights following the inclusion of omitted household members and adjustments to sociodemographic variables • added value labels not yet included in the datafile and applicable to any of the other waves • harmonized variable formats between waves • reworded the labels of selected variables and values • adjusted the user-missing value setting (changed from discrete values to ranges) • corrected data cleaning based on routing rules (especially *eduex2, breakup*) • corrected minor issues in the coding of string variables (e.g. *cpartywv*) • adjusted the width of string variables *hendreo7* to *hendreo10* • coded verbatim answers containing parties into party codes for close-ended options offered from Wave 3 onwards • add variables relating to non-resident children of unused order (6 to 8) • corrected error in the calculation of *nage1* to *nage8* • replaced zeros with system missing values for households with no children aged 10 to 17 years in *h_papi_child*, *h_diary_child* • replaced zeros with the number/share of questionnaires in the household for cases with *rstat* = 5 in *h_indi, hp_indi, h_indi_proxy, hp_indi_proxy, h_papi_adult, hp_papi_adult, h_diary_adult, hp_diary_adult*

The dataset from CHPS Wave 3 was published in a single version on 18 February 2019.

Table 48: Wave 3 dataset versions

	publication date	changes made
v1.0	18 Feb 2019	-

The dataset from CHPS Wave 4 was published in a single version on 21 November 2019.

Table 49: Wave 4 dataset versions

	publication date	changes made
v1.0	21 Nov 2019	-

19. Overview of CHPS documentation for Waves 1 to 4

Table 50: Overview of CHPS documentation for Waves 1 to 4		W1	W2	W3	W4
Advance materials					
Advance letter to the respondents	only in Czech	1	1	1	1
Respondent leaflet	only in Czech	1	1	1	1
Letter to local administrations	only in Czech	1	-	-	-
Letter to city administrations	only in Czech	1	-	-	-
Letter to regional police directorates	only in Czech	1	-	-	-
Business card	only in Czech	1	1	1	1
Thank-you card	only in Czech	1	1	1	1
Data collection instruments					
CAPI household questionnaire		1	1	1	1
CAPI individual questionnaire		1	1	1	1
CAPI proxy questionnaire		1	1	1	1
Adult self-completion *)	MEDIAN version	1	1	1	1
Child self-completion, age 15–17 *)	MEDIAN version	1	-	-	1
Child self-completion, age 10–14 *)	MEDIAN version	1	-	-	1
Child self-completion, age 10–17 *)	MEDIAN version	-	1	1	-
Adult diary		1	1	1	1
Child diary		1	1	1	1
CAPI household questionnaire showcards		1	1	1	1
CAPI individual questionnaire showcards		1	1	1	1
CAPI proxy questionnaire showcards		1	1	1	1
Cognitive ability test record sheet		-	-	-	1
Employment history timeline		-	-	-	1
Dataset					
CAPI and self-completion datasets		1	1	1	1
Child and adult diary datasets		1	1	1	1
Contact form datasets		1	1	1	1
Documentation and miscellaneous					
CAPI and self-completion datasets documentation		1	1	1	1
CAPI and self-completion dataset variable list		1	1	1	1
SPSS syntax for constructing the ISEI and ESeC indicators		1	- **)	- **)	- **)
SPSS syntax for constructing the income variables		1	- **)	- **)	- **)
R script for the weighting		1	1	1	1
Reference distributions used for weighting		1	1	1	1
Child and adult diary data documentation		1	1	1	1
Contact form data documentation		1	1	1	1

*) Variable codes were not included in the questionnaire distributed to the respondents.

**) Wave 2 syntax is equal to that of Wave 1.