

CZECH HOUSEHOLD PANEL SURVEY

Contact Form 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 3.0 (21 November 2019)



Table of Contents

1.	Content of contact forms.....	2
1.1.	Times of contact (<i>month, day, hour, min</i>).....	2
1.2.	Contact outcome (<i>conout</i>).....	2
1.3.	Refusal reasons (<i>conref</i>).....	3
1.4.	Housing type (<i>houstp</i>).....	4
1.5.	Housing quality in the area (<i>haqual</i>).....	4
1.6.	Housing quality comparison (<i>hqualco</i>).....	4
2.	Other variables in the dataset.....	5
3.	Households included in the dataset.....	7
4.	CAWI-surveyed households.....	9
5.	Interviewer data.....	10
6.	Dataset versions.....	11

List of tables

Table 1:	Derived variables and variables from other sources.....	6
Table 2:	Numbers of households in contact form data.....	8
Table 3:	Content of interviewer datasets.....	10
Table 4:	Versions of datasets from contact forms.....	11
Table 5:	Versions of datasets with interviewer information.....	11

1. Content of contact forms

For fieldwork monitoring purposes, the interviewers of the Czech Household Panel Survey (CHPS) completed contact forms in the CAPI administration software to keep records of the circumstances of contacting the households. After every attempt to approach households at selected addresses, the interviewer was to record the *month*, *day*, *hour* and minute (*min*) of contact, contact outcome (*conout*) and any refusal reason (*conref*). The interviewers were also required to determine the type of housing inhabited by the household (*houstp*), assess the quality of housing in the area (*hqual*) and compare the dwelling's quality with that typical of the area (*hqualco*).

1.1. Times of contact (*month, day, hour, min*)

The variables indicate the time at which the contact attempt of specified order took place. The value 99 indicates that other information in the contact form suggests the existence of a record of the interviewer's approaching the household, but the time of contact was not recorded, or the value was lost due to technical problems in the process of storing or transferring data.

The value 95 indicates that the contact attempt of specified order took the form of inviting the household to CAWI, in which the household participated.

In Waves 1 and 2, STEM/MARK interviewers only recorded the month and day of contact, not the hour and minute. Therefore, there are system-missing values in these variables for STEM/MARK households.

1.2. Contact outcome (*conout*)

The interviewers used the following categories to code the outcome of contacting an address:

Contact outcome:	
1	Complete interview obtained
2	Interview refused by household
3	Interview started but incomplete
4	Interview arranged for another date
5	Language barrier – no member speaks Czech
6	Unable to take part (health problems, old age, etc.)
7	No one answered
8	No adult answered
9	Address does not exist/not found
10	Non-residential address

The contact outcome of a specified order is always linked to contact time (*month* to *min*) of the same order. The value 99 indicates that the interviewer failed to record the contact outcome, or the information was lost due to technical problems.

During data processing, the following additional codes were included in the *conout* variables:

11	Household not visited by interviewer	code only kept as final outcome in the <i>outcome</i> variable because the contact form was not completed for non-visited households (i.e. no attempt at personal contact was made at those addresses)
12	Household did not respond to CAWI invite	(from W2) code for non-surveyed households that had taken part in previous wave's CAWI and, therefore, were not approached for face-to-face interviewing
13	Change of address, not interviewed	(from W2) code imputed when interviewer reported relocation of the entire household
14	Interviewed, excluded by quality control	household was interviewed, but subsequent checks revealed failures on the interviewer's part or other surveying errors

1.3. Refusal reasons (*conref*)

When a household refused to participate in the survey, interviewers selected the main refusal reason from the categories below.

What was the main reason for the entire household to refuse participation?	
1	Not enough time (household members constantly busy)
2	Bad timing, extraordinary situation in household (just leaving, illness, visit etc.)
3	No interest/trust in surveys in general
4	Recently responded to a survey
5	Bad prior experience with surveys
6	No interest in this survey's topics
7	No trust in this survey (not enough information, do not like the topic etc.)
8	Concerns about confidentiality/misuse of information/sharing personal data
9	Insufficient incentives – low reward etc.
10	Afraid of interviewer (afraid to open door, afraid of door-to-door selling)
11	A household member “forbade” taking part
12	Other, please complete:
13	No reason given

Refusal reason of specified order is always linked to contact outcome (*conout*) of the same order.

During data processing, the verbal answers to semi-closed question *conref* were assigned to corresponding closed categories. The value 99 indicates that the interviewer failed to record the refusal reason, or the value was lost due to technical problems. When interviewers picked the category 12, and

the follow-up open-ended answer was missing, the *conref* variable was recoded to the value 99. This procedure was chosen because it was possible to code a majority of the open-ended answers given into closed options.

1.4. Housing type (*houstp*)

What type of dwelling does the household reside in?	
1	Farm
2	Detached house
3	Semi-attached house
4	Terraced house
5	Flat in a building with multiple dwelling units
6	Other, please complete:
7	Don't know

1.5. Housing quality in the area (*haqual*)

Please evaluate the quality of buildings/dwellings in this area:	
1	Very good
2	Good
3	Satisfactory
4	Bad
5	Very bad

1.6. Housing quality comparison (*hqualco*)

Please evaluate the quality of the household's dwelling compared to the quality of buildings/dwellings in this area:	
1	Considerably higher
2	Somewhat higher
3	The same
4	Somewhat lower
5	Considerably lower

The values 99 in the *houstp*, *haqual* and *hqualco* variables indicate that the interviewer failed to record the answer, or the value was lost due to technical problems.

In Wave 2, only STEM/MARK interviewers inquired about this information, while system-missing values were imputed for MEDIAN households of that wave.

From Wave 2, the households contacted through CAWI only obtained the value 95 (CAWI only) in these variables.

During data processing, verbal answers to the half-closed question *houstp* were assigned to corresponding closed categories.

2. Other variables in the dataset

The datasets include household ID variables (*hid*) for linking to other types of data. Current and previous waves' IDs are included.

During data processing, a few more variables were added to the datasets that were derived from contact form questions or other sources of interviewing context information.

The *coform* variable indicates whether contact form answers (*month* – *hqualco*) are available for the household. In some cases, the interviewers failed to complete the contact forms, or the data was lost due to technical problems. Households without contact form data were assigned the value 0 in *coform* and system-missing values in the variables *month* to *hqualco*. Households for which last wave's CAWI was continued always have their *coform* equal to 1.

The *result* variable indicates whether the household was interviewed in the given wave (1) or not (0). The value 2 was imputed when the household was interviewed but subsequent controls revealed failures on the interviewer's part or other problems (e.g., the household CAPI was not finished, or basic sociodemographics about self-completion respondents were inconsistent with CAPI data). These households' *outcome* variable equals 14.

The *agen* variable indicates the fieldwork agency that was supposed to survey the household.

The *qmode* variable indicates the data collection mode, distinguishing between CAPI (1) and CAWI (2). Households that were not surveyed were coded as 0.

The *convert* variable equals one for households that were invited to CAWI after refusing CAPI or unsuccessful contact attempts by the interviewer. This variable is available from Wave 2 because CAWI conversion in Wave 1 was not recorded systematically at the household level.¹ From Wave 2, MEDIAN offered CAWI as a conversion strategy *en masse*, especially in major cities, whereas STEM/MARK only offered it in exceptional cases.

The *intid* variable identifies the interviewer who surveyed the household. The assigned interviewer's ID is indicated for households that were converted to CAWI or not surveyed. Households surveyed by CAWI in a previous wave were sent a mail or e-mail invite to CAWI without being approached by the interviewer. In these cases, the *intid* variable is coded as 999,996. Households in areas in which no interviewer could be deployed are coded as 999,998.

The final contact outcome is indicated by the *outcome* variable, which is derived from the *conout* variables. For households for which no contact form was available (*coform* = 0) and which were not interviewed (*result* = 0), the *outcome* variable was set as "household not visited" (11). When the household was visited (*result* = 1), the *outcome* variable was set as "complete interview obtained" (1).

Territorial identification of the address is given in the variables *c_msize* (size of municipality of residence), *c_region* (region) and *c_county* (county). Wave 1 data refer to address at the time of

¹ In the main dataset's documentation, the CAWI conversion rate for Wave 1 was estimated from the total numbers of refusals and non-contacts.

sampling, the data for subsequent waves refer to the place of residence during previous wave's interview.

Wave 1 dataset includes the variables *strata* and *psu* that identify the household's stratum during stratified sampling and the primary sampling unit, respectively.

Table 1: Derived variables and variables from other sources		Wave
w(x)_hid	unique household ID in Wave (x)	W1 +
coform	whether contact form information is available for the household	W1 +
result	whether household was interviewed successfully (1), not interviewed (0) or interviewed but excluded from dataset (2)	W1 +
agen	fieldwork agency	W1 +
qmode	whether household took part in personal (1) or CAWI (2) interviewing or was not surveyed (0)	W1 +
convert	CAPI to CAWI conversion attempted (1) or not (0)	W2 +
intid	interviewer ID (the same interviewer has the same ID across survey waves)	W1 +
outcome	final contact outcome	W1 +
refusal	final refusal reason	W1 +
c_msize	household municipality size during sampling/last wave's interviewing	W1 +
c_region	household region during sampling/last wave's interviewing	W1 +
c_county	household county during sampling/last wave's interviewing	W1 +
strata	stratum (1 to 58)	W1
psu	primary sampling unit (1 to 1,275)	W1

3. Households included in the dataset

Wave 1 dataset contains information on **12,810** households. In addition to 12,750 households from the sample, it includes 60 successfully surveyed households that represent second and further households on common budget found in the dwellings sampled (when the dwelling was inhabited by more than one household on common budget).

Any further households that were not surveyed are not indicated in contact form data. Those households were not recorded systematically, and their number was derived from the *ohous* variable in household CAPI data (assuming there is only one more household in the dwelling). In the main dataset's documentation, the non-surveyed further households in a dwelling are included under the "household not visited" category in the field outcomes overview.

During Wave 1 fieldwork, 312 address deemed non-residential were substituted for additional addresses from the same primary sampling unit in the gross sample. The new addresses are not included side-by-side with the original addresses but rather replace them in the contact form dataset.

Wave 2 dataset consists of **5,109** households, including 5,093 households surveyed in Wave 1 (of the 5,159 surveyed, 67 CAWI households were not issued because their contact information was incorrect or not available; in contrast, one household excluded from the Wave 1 dataset due to interviewer failure was issued and successfully interviewed) and 16 split-off households.

Wave 3 dataset comprises **4,158** households, including 4,147 households surveyed in Wave 2 and 11 split-off households. **Wave 4** dataset consists of **3,626** households, including 3,616 households surveyed in Wave 3 and 10 split-off households.

The second and other parts of split households were only included in the contact form data when they were interviewed. Those further parts of split households that were not surveyed are not recorded systematically in the contact forms. In the main dataset's documentation, their number was estimated from the *rstat* variable (members' residential status) in the household CAPI, assuming that a household with non-deceased leavers split in two. In the main dataset's documentation, the non-surveyed second halves of the split households are included under the "household not visited" category in the field outcomes overview.

Table 2 summarizes the numbers of households included in the contact form dataset and deemed issued in the main dataset's documentation. The *contact outcome* column indicates the type of contact outcome assigned by the main dataset's documentation for the estimated numbers of non-surveyed further households in a dwelling and split-off households, as well as for the substitute households for Wave 1's non-residential addresses. To keep the contact outcome statistic simple, all non-surveyed parts of split households are included under category 13 (change of address, not interviewed) even if some of them did not relocate to a new address².

² Estimating on the basis of the *rstat* and *asame* variables, 95% of non-surveyed parts of split households relocated.

Table 2: Numbers of households in contact form data

	W1	W2	W3	W4	Contact outcome
net sample	12,750	-	-	-	-
further household in dwelling, surveyed	60	-	-	-	-
further household in dwelling, not surveyed *)	147	-	-	-	11 (not visited)
substitute addresses	312	-	-	-	10 (non-residential)
surveyed in previous wave (<i>result</i> = 1)	-	5,092	4,147	3,616	-
excluded in previous wave (<i>result</i> = 2)	-	1	0	0	-
split-off households , surveyed	-	16	11	10	-
split-off households , non-surveyed *)	-	155	145	108	13 (change of address)
contact form addresses, TOTAL	12,810	5,109	4,158	3,626	
addresses issued, TOTAL	13,269	5,264	4,303	3,734	

*) Estimate based on household CAPI.

4. CAWI-surveyed households

In Wave 1, households refusing personal participation and repeated non-contacts were sent a letter of invitation to complete an online questionnaire, including a unique access code to the CAWI mutation of the survey. For households that agreed to CAWI, the outcome of the contact (*conout*) following the last personal contact is coded as “complete interview obtained” (1) and the contact times (*month to min*) are coded as 95. Online interviewing is always accounted for as the last contact, even if the household may actually have been approached following CAWI completion by an interviewer who had not been informed of their CAWI in time.

For unsuccessful CAWI conversions, the CAWI invitation is not recorded in the *month to min* and *conout* variables, and the final contact outcome (*outcome*) is assigned one of the interviewer-selected codes. In Wave 1, CAWI invites were not recorded systematically, and therefore, the contact form dataset does not indicate the CAWI conversion attempts.

From Wave 2, the addresses of previous-wave CAWI households were not issued to interviewers for face-to-face interviewing. Instead, a repeated invite to complete an online questionnaire was sent to them by mail or e-mail. The first-order contact outcome for these households (*conout*) is coded either as “complete interview obtained” (1) or “did not respond to CAWI invite” (12). First-order contact times are coded as 95.

From Wave 2, households for which a CAWI conversion attempt was made were marked with value 1 in the *convert* variable. To preserve consistency with Wave 1 dataset, unsuccessful CAWI conversions are not recorded as last-order contacts in the *month to cmin* and *conout* variables.

From Wave 2, households that were invited directly to CAWI and whose addresses were, therefore, not issued to interviewers for face-to-face interviewing, the variables *houstp*, *haqual* and *hqualco* were coded as 95.

5. Interviewer data

The questionnaire and contact form datasets contain a unique interviewer ID (*intid*) that remains the same throughout all the waves in which the interviewer worked. The *intid* variable links the datasets with basic information on the interviewers.

Basic information on interviewers is available in separate files prepared for each wave. These characteristics include sex (*insex*), birth year (*inbirthy*), educational attainment (*ineduc*), and a categorized indicator of length of experience with the agency (*inexp*). The educational attainment and length of experience variables refer to the situation at the start of the wave's fieldwork.

The agencies' internal interviewer IDs were replaced with new random IDs in the datasets.

The interviewer datasets include both interviewers who contributed at least one interviewed household to the survey and those who did not survey any of the households issued to them.

Table 3: Content of interviewer datasets

intid	interviewer ID (the same interviewer has the same ID across survey waves)
agen	fieldwork agency
insex	interviewer sex
inbirthy	interviewer year of birth
ineduc	interviewer educational attainment
inexp	interviewer experience with the agency

6. Dataset versions

By 18 February 2019, the datasets from CHPS Wave 1 and 2 contact forms had been published in two versions. In versions 1.1, the variable *c_msize* was adjusted so that the classification of municipality size is based on the same administrative data as the classification in the main data files. Territorial information was included in Wave 2 new version about a household that had been excluded from Wave 1 data due to interviewer misconduct but successfully interviewed in Wave 2 and verified.

The dataset from CHPS Wave 3 contact forms was published in a single version on 18 February 2019. The dataset from CHPS Wave 4 contact forms was published in a single version on 21 November 2019.

The data files with interviewer information from Waves 1 to 4 have been published in a single version.

Table 4: Versions of datasets from contact forms

	Wave 1	Wave 2	Wave 3	Wave 4
v1.0	20 Feb 2018	20 Feb 2018	18 Feb 2019	21 Nov 2019
v1.1	18 Feb 2019	18 Feb 2019	-	-

Table 5: Versions of datasets with interviewer information

	Wave 1	Wave 2	Wave 3	Wave 4
v1.0	20 Feb 2018	20 Feb 2018	18 Feb 2019	21 Nov 2019