

ČESKÉ PANELOVÉ ŠETŘENÍ DOMÁCNOSTÍ

Dokumentace k datům

Vlna 1 (2015) – vlna 4 (2018)

Mgr. Michaela Kudrnáčová

Sociologický ústav AV ČR, v. v. i.

www.promenyceskespolecnosti.cz

promeny.spolecnosti@soc.cas.cz

Verze 5.0 (21.11. 2019)



Obsah

Shrnutí	5
1.	České panelové šetření domácností.....7
2.	Výběrový soubor9
2.1.	Cílová skupina..... 9
2.2.	Jednotka výběru 9
2.3.	Výběr vzorku domácností..... 10
2.4.	Sledování členů panelu („following rules“) 11
3.	Sběr dat 13
3.1.	Školení tazatelů 13
3.2.	Pilotní šetření 14
3.3.	Avizování výzkumu 14
3.4.	Průběh terénních prací..... 15
3.5.	Kontaktování domácností..... 19
3.6.	CAWI dotazování 21
3.7.	Monitorování terénních prací 21
3.8.	Rozhovor v domácnosti 22
3.9.	Odměny pro respondenty a tazatele..... 22
3.10.	Kontrola práce tazatele 23
3.11.	Udržování kontaktu s domácnostmi..... 23
4.	Dotazové instrumenty 24
4.1.	Typy instrumentů 24
4.2.	Distribuce instrumentů..... 25
4.3.	Návratnost na úrovni instrumentů..... 25
4.4.	Dokumentace instrumentů 30
4.5.	CAWI mutace dotazových instrumentů 34
4.6.	Konzistence kódů odpovědí mezi vlnami 34
4.7.	Proměnné opakované v cyklech..... 36
4.8.	Podmíněné dotazování..... 36
4.9.	Kognitivní testy (vlna 2) 37
4.10.	Zaměstnanecká historie (vlna 2) 38
5.	Obsah a struktura datového souboru 39
5.1.	Řazení proměnných 39

5.2.	Struktura souboru	40
5.3.	Proměnné nezařazené do datového souboru	40
6.	Identifikační proměnné	43
7.	Proměnné popisující změnu složení domácnosti	45
8.	Proměnné v podmíněném dotazování	47
9.	Generované proměnné	50
10.	Stratifikační postavení	55
11.	Proměnné individuálního příjmu	56
12.	Proměnné vztahující se k okolnostem dotazování	59
13.	Proměnné filtrované na základě hodnot jiné osoby	62
14.	Anonymizace dat	64
15.	Kontrola, zpracování a čištění dat	65
15.1.	Kontroly dat	65
15.2.	Vyřazování dotazníků	66
15.3.	Tvorba datového souboru	66
15.4.	Čištění dat	66
15.5.	Kódování otevřených a polootevřených otázek	68
15.6.	Imputace hodnot	69
15.7.	Úpravy počtu případů podle navazujících vln	69
15.8.	Neplatné odpovědi/hodnoty	69
16.	Specifika vzorku	72
17.	Post-stratifikační váhy	73
17.1.	Binární proměnné určující vyplnění instrumentů	73
17.2.	Metoda vážení	75
17.3.	Proměnné pro vážení	75
17.4.	Práce s chybějícími hodnotami	76
18.	Verze datového souboru	78
19.	Přehled materiálů k 1. až 4. ročníku CHPS	81

Přehled tabulek

Tabulka 1: tematické okruhy Českého panelového šetření domácností (CHPS).....	7
Tabulka 2: věkové vymezení populace jednotlivých instrumentů	9
Tabulka 3: výsledky kontaktů ve vzorku domácností, vlna 1	15
Tabulka 4: výsledky kontaktů ve vzorku domácností, vlna 2	16
Tabulka 5: výsledky kontaktů ve vzorku domácností, vlna 3	17
Tabulka 6: výsledky kontaktů ve vzorku domácností, vlna 4	18
Tabulka 7: důvody odmítnutí účasti ve výzkumu	20
Tabulka 8: počty vyšetřených domácností a míry retence a konverze podle způsobu sběru dat	21
Tabulka 9: průměrná délka CAPI dotazníků v minutách	24
Tabulka 10: počty vyplněných instrumentů, vlna 1	26
Tabulka 11: míra vyplnění instrumentů celkem a podle typu domácnosti (hhtype), vlna 1.....	26
Tabulka 12: počty vyplněných instrumentů, vlna 2	27
Tabulka 13: míra vyplnění instrumentů celkem a podle typu domácnosti (hhtype), vlna 2.....	27
Tabulka 14: počty vyplněných instrumentů, vlna 3	28
Tabulka 15: míra vyplnění instrumentů celkem a podle typu domácnosti (hhtype), vlna 3.....	28
Tabulka 16: počty vyplněných instrumentů, vlna 4	29
Tabulka 17: míra vyplnění instrumentů celkem a podle typu domácnosti (hhtype), vlna 4.....	29
Tabulka 18: proměnné s odlišnostmi v kódování odpovědí.....	35
Tabulka 19: pořadí skupin proměnných v datovém souboru.....	39
Tabulka 20: struktura datového souboru.....	40
Tabulka 21: proměnné nezařazené do datového souboru	41
Tabulka 22: identifikační proměnné jednotlivce/domácnosti	43
Tabulka 23: identifikační čísla dalších členů domácnosti	44
Tabulka 24: proměnné vztahující se k novým a bývalým členům domácnosti	46
Tabulka 25: počty nových a bývalých členů domácností	46
Tabulka 26: proměnné vztahující se k podmíněnému dotazování.....	49
Tabulka 27: generované proměnné na úrovni domácnosti (CAPI dotazník pro domácnost)	50
Tabulka 28: ukazatele přítomnosti skupin osob v domácnosti	51
Tabulka 29: typ domácnosti podle struktury	52
Tabulka 30: generované proměnné na úrovni jednotlivce	53
Tabulka 31: ukazatele stratifikačního postavení	55
Tabulka 32: způsob konstrukce proměnných dignum a dinnum	57

Tabulka 33: proměnné individuálního příjmu	58
Tabulka 34: místo bydliště domácnosti.....	59
Tabulka 35: charakteristiky dotazování.....	60
Tabulka 36: datum dotazování.....	60
Tabulka 37: vyplnění dotazníků v domácnosti	61
Tabulka 38: proměnné filtrované na základě hodnot jiné osoby.....	62
Tabulka 39: čištění proměnných dayreg až daylifm	68
Tabulka 40: neplatné odpovědi a jejich kódy v CAPI a PAPI datech.....	70
Tabulka 41: CAPI proměnné s chybějícími hodnotami v důsledku problémů scriptu.....	71
Tabulka 42: indikátory skupin respondentů a příslušné váhy	73
Tabulka 43: skupiny proměnných pro vážení.....	75
Tabulka 44: kategorie proměnných pro vážení.....	76
Tabulka 45: přehled dílčích vzorků respondentů a jejich vážení.....	77
Tabulka 46: verze datového souboru z 1. ročníku	78
Tabulka 47: verze datového souboru z 2. ročníku	79
Tabulka 48: verze datového souboru z 3. ročníku	80
Tabulka 49: verze datového souboru z 3. ročníku	80
Tabulka 50: přehled materiálů k 1. až 4. ročníku CHPS.....	81

Přehled schémat

Schéma 1: ilustrace pravidel pro sledování jednotlivců a domácností.....	12
Schéma 2: popis základních náležitostí publikovaných CAPI dotazníků	30
Schéma 3: další náležitosti publikovaných CAPI dotazníků.....	31
Schéma 4: rozlišení parametrů otázek pro původní a nové respondenty v CAPI INDI dotazníku	32
Schéma 5: srovnání původní a publikované podoby PAPI dotazníků	33

Shrnutí

Cílová skupina

Neinstitucionalizovaná populace České republiky

Jednotka výběru

Hospodařící domácnost

Předmět šetření

V prvním roce výzkumu všichni členové vybraných domácností. Ve druhém až čtvrtém roce všichni členové domácností, v nichž žil alespoň jeden člen panelu z předchozího roku.

Výběr vzorku

Dvoustupňový stratifikovaný náhodný výběr

Realizátor sběru dat

MEDIAN, s.r.o., a STEM/MARK, a.s.

Termín sběru dat

vlna 1	vlna 2	vlna 3	vlna 4
7. 7. – 1. 12. 2015	20. 6. – 31. 10. 2016	21. 6. – 30. 10. 2017	20. 6 – 15. 10. 2018

Počet vyšetřených domácností

vlna 1	vlna 2	vlna 3	vlna 4
5 159	4 147	3 616	3 188

Návratnost a retence na úrovni domácností

vlna 1	vlna 2	vlna 3	vlna 4
42,1 %	79,5 %	84,7 %	86,4 %

Způsob sběru dat

Standardizovaný rozhovor face-to-face s využitím počítače („computer assisted personal interviewing“, CAPI), vyplňování standardizovaného dotazníku na papíře („paper-and-pencil self-administered questionnaire“, SAQ), vyplňování standardizovaného dotazníku na internetu („computer assisted web interviewing“, CAWI)

Jednotka analýzy

Domácnost a jednotlivec

Obsah a struktura datového souboru

Hlavní datový soubor, k němuž se váže tato dokumentace, obsahuje proměnné z následujících instrumentů: CAPI dotazník pro domácnost, CAPI individuální a proxy dotazník, PAPI dotazník

pro dospělé, PAPI dotazník pro děti 10–14 let a 15–17 let, včetně identifikačních a odvozených proměnných, proměnných vztahujících se k průběhu dotazování a vah.

Data z časových snímků a data z kontaktních formulářů jsou dostupná v samostatných souborech s vlastní dokumentací. Všechny typy datových souborů jsou pořízeny zvlášť za jednotlivé vlny výzkumu.

Upozornění k označování dotazníků pro samostatné vyplnění na papíře

Dotazníky pro samostatné vyplnění na papíře jsou v dokumentaci, názvech dotazových instrumentů i v datových souborech označovány jako PAPI dotazníky, ačkoli nebyly určeny pro rozhovor s tazatelem. Toto nepřesné označení bylo zavedeno během sběru dat, protože umožnilo jednoduché a snadno zapamatovatelné rozlišení tří základních typů dotazových instrumentů: CAPI dotazník (pro domácnost, jednotlivce a proxy), dotazník pro samovyplnění (pro děti a dospělé) a časový snímek pro samovyplnění (pro děti a dospělé). Tyto tři typy instrumentů představují kombinace módu sběru dat (CAPI rozhovor vs. samostatné vyplnění dotazníku na papíře) a formátu dotazníku (běžný dotazník s převážně uzavřenými otázkami vs. časový snímek). Výraz „PAPI dotazník“, chápán v obecném významu jako „dotazník na papíře“, se v tomto kontextu ukázal jako stručné a srozumitelné označení umožňující odlišit dotazník pro samostatné vyplnění na papíře od dotazníku pro CAPI rozhovor na jedné straně a časového snímku na straně druhé.

Kontrola a čištění dat

Kontrola rozsahu hodnot a filtrování, čištění základních socio-demografických proměnných, kódování (polo)otevřených otázek

Vážení dat

Post-stratifikační váhy k 16 dílčím vzorkům respondentů založené na kraji bydliště, pohlaví, věku a vzdělání. U respondentů časových snímků jsou mimo základních vah dostupné i varianty pro zajištění uniformní distribuce dnů v týdnu.

1. České panelové šetření domácností

České panelové šetření domácností (CHPS) je výběrové šetření, v němž byl v letech 2015–2018 opakovaně dotazován náhodně vybraný vzorek domácností žijících na území České republiky. V první vlně sběru dat, která proběhla od července do prosince 2015, bylo dotázáno 5 159 domácností. V následujících třech letech byly o účast požádány ty domácnosti, které poskytly odpovědi předchozí rok. Závěrečné, čtvrté, vlny výzkumu se v roce 2018 zúčastnilo 3 188 domácností.

Pro komunikaci s veřejností je pro výzkum využíván název Proměny české společnosti (www.promenyceskespolecnosti.cz).

Cílem výzkumu je zmapovat životní podmínky českých domácností z dlouhodobé perspektivy, popsat dynamiku změny v životě domácností a jednotlivců a vztáhnout proces sociální změny ke vztahům a dění v domácnostech.

CHPS je mezioborový výzkum využívající přístupy sociologie, ekonomie a politologie. Zahrnuje pět hlavních tematických okruhů (*tabulka 1*).

Tabulka 1: tematické okruhy Českého panelového šetření domácností (CHPS)

Rodinný život, užívání času, zdraví	Vzdělání a trh práce	Sociální stratifikace	Bydlení	Politická participace a občanská společnost
<ul style="list-style-type: none">• Vztahy v domácnostech	<ul style="list-style-type: none">• Vzdelanostní dráhy a šance	<ul style="list-style-type: none">• Sociální nerovnosti	<ul style="list-style-type: none">• Charakteristiky bydlení	<ul style="list-style-type: none">• Volební účast a preference
<ul style="list-style-type: none">• Sladění práce a rodiny	<ul style="list-style-type: none">• Vzdelanostní aspirace	<ul style="list-style-type: none">• Třídní struktura• Sociální mobilita	<ul style="list-style-type: none">• Spokojenost s bydlením	<ul style="list-style-type: none">• Stranické sympatie
<ul style="list-style-type: none">• Péče o děti	<ul style="list-style-type: none">• Zaměstnanost, nezaměstnanost	<ul style="list-style-type: none">• Příjmy a bohatství	<ul style="list-style-type: none">• Preferované bydlení	<ul style="list-style-type: none">• Politické a občanské postoje
<ul style="list-style-type: none">• Hospodaření s příjmy	<ul style="list-style-type: none">• Pohyb na pracovním trhu	<ul style="list-style-type: none">• Kulturní kapitál	<ul style="list-style-type: none">• Dostupnost bydlení	<ul style="list-style-type: none">• Občanská angažovanost
<ul style="list-style-type: none">• Genderové postoje	<ul style="list-style-type: none">• Pracovní podmínky		<ul style="list-style-type: none">• Mezigenerační přenos vlastnictví a postojů k bydlení	<ul style="list-style-type: none">• Diskuse o politice v rodinách
<ul style="list-style-type: none">• Užívání času				
<ul style="list-style-type: none">• Zdraví a duševní pohoda	<ul style="list-style-type: none">• Spokojenost s prací			

Výzkum organizoval Sociologický ústav Akademie věd ČR, v.v.i., CERGE-EI (společné pracoviště Centra pro ekonomický výzkum a doktorské studium Univerzity Karlovy a Národohospodářského ústavu Akademie věd ČR) a Fakulta sociálních studií Masarykovy univerzity. Výzkumný tým složený ze členů těchto institucí vytvořil metodický plán výzkumu a dotazové instrumenty, vybral ve veřejné soutěži realizátory terénních prací, s realizátory terénních prací spolupracoval na implementaci metodiky výzkumu a kontrole a čištění dat, zpracoval konečné datové soubory a zajistil jejich uložení v Českém sociálně vědním datovém archivu Sociologického ústavu AV ČR, v.v.i.

Sběr dat realizovaly agentury MEDIAN, s.r.o., a STEM/MARK, a.s., které patří k nejvýznamnějším agenturám pro výzkum trhu a veřejného mínění v České republice. Jsou zakládající členové organizace SIMAR (Sdružení agentur pro výzkum trhu a veřejného mínění) a členové ESOMAR (European Society for Opinion and Marketing Research). Mimo sběr dat provedli realizátoři výběr vzorku, spolupracovali na kontrole a čištění dat a poskytovali konzultace k metodice výzkumu.

Výzkum finančně podpořila Grantová agentura ČR z grantu GB14-36154G (Dynamika změny v české společnosti).

2. Výběrový soubor

2.1. Cílová skupina

Cílovou skupinu šetření tvoří neinstytucionalizované obyvatelstvo České republiky bez ohledu na občanství a národnost. Kritériem pro zařazení domácnosti do výzkumu byla schopnost zúčastnit se dotazování, tj. alespoň někteří členové domácnosti hovořili česky nebo slovensky a v účasti v dotazování jim nebránil zdravotní stav.

Věkové vymezení populace se liší podle jednotlivých dotazových instrumentů, které byly určeny různým věkovým skupinám členů domácnosti (*tabulka 2*).

Tabulka 2: věkové vymezení populace jednotlivých instrumentů

CAPI dotazník pro domácnost	bez věkového omezení
CAPI individuální dotazník	18+ let
CAPI proxy dotazník	18+ let
PAPI dotazník pro dospělé	18+ let
PAPI dotazník pro starší děti	15–17 let
PAPI dotazník pro mladší děti	10–14 let
časový snímek pro dospělé	18+ let
časový snímek pro děti	10–17 let

2.2. Jednotka výběru

Jednotkou výběru je hospodařící domácnost definovaná v souladu s metodikou Českého statistického úřadu.

Domácnost tvoří jedinci obvykle bydlící v daném bytě či rodinném domku, kteří společně hospodaří, tj. společně hradí základní nezbytné výdaje domácnosti, jako je strava, náklady na bydlení, ostatní provozní výdaje domácnosti (včetně těch osob, které na náklady samy nepřispívají, ale hradí jim je další členové domácnosti, např. děti, osoby bez příjmu).

Předmětem šetření jsou všichni členové vybraných domácností. Pro určení členství v domácnosti je rozhodující obvyklý pobyt, nikoli trvalý pobyt.

Mezi členy domácnosti patří (pokud společně hospodaří):

- 1 osoby, které v bytě obvykle bydlí (zdržují se zde většinu mimopracovního/mimoškolního času, a to během minulých 6 měsíců, případně se v bytě plánují zdržovat alespoň 6 měsíců);
- 2 podnájemníci a hosté, kteří v současnosti nesídlí na jiné soukromé adrese anebo jejichž zamýšlená délka pobytu v šetřeném bytě je delší než 6 měsíců;
- 3 osoby dlouhodobě studující a pracující mimo domov nezávisle na délce své nepřítomnosti, pokud nemají žádnou jinou soukromou adresu a udržují úzkou vazbu na šetřenou domácnost;
- 4 osoby dlouhodobě pobývajících v nemocnici a dalších institucích, pokud délka jejich nepřítomnosti nepřekročí 6 měsíců a udržují úzkou finanční vazbu na šetřenou domácnost.

Studenti dlouhodobě pobývajících mimo rodičovskou domácnost jsou považováni za členy rodičovské domácnosti, pokud nemají žádnou další soukromou adresu, na níž se zdržují častěji než v domácnosti rodičů, a mají finanční vazbu na rodičovskou domácnost (v případě pochybnosti bylo aplikováno rozhodovací pravidlo, zda domácnost dala v posledním roce studentovi alespoň 12 000 Kč).

2.3. Výběr vzorku domácností

Vzorek byl získán stratifikovaným dvoustupňovým náhodným výběrem, který je však díky svému designu srovnatelný se stratifikovaným jednostupňovým skupinkovým výběrem („cluster sampling“). Efekt designu je dále redukován využitím velkého počtu malých výběrových bodů („primary sampling unit“).

Datový soubor obsahuje proměnné *strata* a *psu*, které umožňují korekci standardních chyb počítaných za předpokladu prostého náhodného výběru.

Výběr vzorku provedly agentury MEDIAN, s.r.o, a STEM/MARK, a.s. Jako opora výběru byl využit aktuální Registr sčítacích obvodů a budov¹, který byl pro účely výběru převeden do formy databáze adres (zahrnutí budovy do databáze tolikrát, kolik obsahuje bytů). Stupně výběru:

- 1 Území ČR bylo rozděleno na 58 strat vymezených jednotkami NUTS 3 (kraje) a 5 kategoriemi velikosti sídla (z teoretických 70 strat neobsahovalo 12 strat žádná sídla). Pomocí systematického náhodného výběru byl v každém stratu zvolen takový počet výběrových bodů (adres), aby odpovídal počtu zde žijících domácností (podle dat Sčítání lidu, domů a bytů 2011, Český statistický úřad) a předpokládané návratnosti. Celkový počet výběrových bodů byl 1 275. Před provedením systematického náhodného výběru byly jednotky databáze seřazeny podle atributů, jako je číslo územní jednotky nebo číslo budovy.
- 2 K adresám vymežujícím výběrové body bylo na základě geografické blízkosti prostým náhodným výběrem přiřazeno dalších 16 adres. V obcích s více než 999 obyvateli byly vybírány adresy vzdálené ne více než 500 metrů od primární adresy. V obcích s maximálně 999 obyvateli byly další adresy vybírány v rámci dané obce bez omezení vzdálenosti od primární adresy. Výsledný hrubý výběrový soubor čítal 21 675 adres (1 274 * 17).

¹ Registr sčítacích obvodů a budov je veřejný seznam, v němž jsou evidovány budovy a jejich atributy. Seznam vede Český statistický úřad. Více informací na https://www.czso.cz/csu/rso/registr_scitacich_obvodu

Adresy z výběrového souboru byly následně vydány pracovníkům agentur (nikoli tazatelům provádějícím sběr dat), kteří ověřovali existenci adres a jejich bytový charakter. Neexistující a nebytové adresy vyřadili a zapsali jména uvedená na schránkách, zvoncích atd. Cílem tohoto kroku bylo omezit počet neexistujících a nebytových adres v hlavním sběru a umožnit personalizované avizování výzkumu. V ověřeném hrubém souboru bylo v každém výběrovém bodě náhodně zvoleno 10 adres. Výsledný čistý výběrový soubor obsahoval 12 750 adres.

Během terénních prací prvního ročníku byla provedena substituce 312 adres, které se ukázaly být neobydlené nebo dlouhodobě neobývané. Místo těchto adres byly náhodně vybrány nevyužité adresy z hrubého výběrového souboru ve stejných výběrových bodech.

Pokud se v bytě nacházelo více hospodařících domácností, nebyl stanoven žádný postup pro výběr některé (některých) z nich, ale tazatelé měli pokyn pokusit se dotázat všechny.

2.4. Sledování členů panelu („following rules“)

Ve druhé až čtvrté vlně šetření spadaly do výběrového souboru takové domácnosti, v nichž v době sběru dat žila alespoň jedna osoba, která byla členem domácnosti dotázané v bezprostředně předcházejícím ročníku („original sample member“).² Pokud tedy domácnost nebyla ve druhém roce dotázána, ve třetím a čtvrtém roce již její členové k účasti přizváni nebyli a totéž platí pro ukončení účasti ve třetí vlně.

Šetřeny byly pouze domácnosti pobývající mimo instituce na území České republiky. Pokud se tedy někteří nebo všichni členové domácnosti po první vlně přesunuli trvale do institucionálního zařízení nebo vystěhovali z České republiky, nebyli dále za účelem dotazování sledováni. Byli však započítáni jako členové svých původních domácností v tom případě, že se k nim v dalších letech vrátili.

Ve druhém až čtvrtém ročníku byly předmětem šetření všechny osoby ve výše vymezených domácnostech, tedy i noví členové, kteří v minulých letech k vzorku nepatřili. Noví členové domácnosti získávali v dalších letech status původních členů domácnosti. Byli tedy sledováni až do konce výzkumu, nikoli jen do doby, kdy žili v domácnostech původních členů z prvního roku. Status dočasných členů domácnosti („temporary sample member“), existující v některých jiných panelových výzkumech domácností, nebyl v CHPS využit.

Při přestěhování celé domácnosti na novou adresu měla být tato domácnost dotázána v místě svého nového bydliště. Nahrazení přestěhované domácnosti novou domácností v dané bytové jednotce nebylo možné. Pokud rozdělením původní domácnosti vzniklo více nových domácností (např. odstěhování dospělého dítěte od rodičů a založení samostatné domácnosti), bylo cílem dotázat jak domácnost na původní adrese, tak odvozenou domácnost na nové adrese. Všechny nově vzniklé domácnosti měly být vyšetřeny i v případě, kdy na původní adrese nezůstala žádná z nich (tj. všechny dceřiné domácnosti se přesunuly na jinou adresu), nebo naopak k rozdělení došlo v rámci původní

² Ve druhém roce výzkumu byly vyšetřeny 2 domácnosti, kde aktuálně nežila žádná osoba z předchozího roku, ale v době od minulého dotazování byla součástí této domácnosti. K původním členům se v obou domácnostech v mezidobí přistěhovali noví členové, avšak původní členové před začátkem sběru dat zemřeli.

adresy (např. rozdělení původní vícegenerační domácnosti na dvě samostatně hospodařící domácnosti v jednom bytě).

Následující dva příklady shrnuté ve *schématu 1* ilustrují princip pravidel pro sledování domácností a jednotlivců.

Příklad A: V první vlně byli dotázáni manželé Adéla a Adalbert. Před druhou vlnou přišli o vlastní bydlení a přestěhovali se k dceři Agátě. Ve druhé vlně byla tedy vyšetřena domácnost Adély, Adalberta a Agáty. Ve třetí vlně se Adéla a Adalbert odstěhovali od Agáty, ale nadále se odmítli výzkumu zúčastnit. K Agátě se nastěhoval přítel Alex a jejich domácnost byla dotázána. Ve čtvrtém roce nebyli Adéla a Adalbert osloveni, protože se nezúčastnili třetí rok. Do domácnosti Agáty a Alexe se narodila dcera Aneta a tato domácnost se opět zúčastnila.

Příklad B: První rok byla dotázána domácnost, v níž žili manželé Bořek a Beáta a jejich syn Boris. Ve druhé vlně se Boris přestěhoval k přítelkyni Betty a jejich nová domácnost se zapojila do výzkumu. Zúčastnila se také domácnost nyní už jen Bořka a Beáty. Třetí rok pokračovala ve výzkumu jak domácnost Bořka a Beáty, tak domácnost Borise a Betty, k nimž se nastěhovala Borisova sestra Bára. Bořek s Beátou se čtvrtý rok již účastnit nechtěli. Z domácnosti Borise a Betty se odstěhovala Bára. K účasti byla tedy oslovena jak domácnost Borise a Betty, tak nová jednočlenná domácnost Báry. Obě tyto domácnosti se čtvrtý rok zúčastnily.

Schéma 1: ilustrace pravidel pro sledování jednotlivců a domácností

vlna 1	vlna 2	vlna 3	vlna 4
Adéla Adalbert	Adéla Adalbert Agáta	odmítnutí Agáta Alex	neosloveni Agáta Alex Aneta
Bořek Beáta Boris	Bořek Beáta Boris Betty	Bořek Beáta Boris Betty Bára	odmítnutí Boris Betty Bára

3. Sběr dat

3.1. Školení tazatelů

Pro práci na CHPS byli vybíráni především tazatelé s dlouholetou praxí práce pro danou agenturu a se zkušeností s výzkumem na náhodně vybraném vzorku a CAPI dotazováním. Všichni tazatelé pracující na CHPS prošli v každém roce speciálním školením.

V průběhu června 2015 probíhala ve vybraných městech České republiky hromadná jednodenní školení pro první ročník výzkumu. Několik tazatelů bylo doškoleny individuálně. Před sběrem dat ve druhém až čtvrtém ročníku absolvovalo přibližně 80 % tazatelů jednodenní hromadné školení během května a června příslušného roku a zbylí tazatelé byli doškoleni individuálně (buď osobně nebo pomocí e-learningového kurzu).

Školení tazatelů se ve všech letech zaměřovalo na:

- význam a cíl výzkumu, akademické instituce organizující výzkum,
- časový plán výzkumu v letech 2015–2018,
- postup výběru vzorku domácností,
- definice domácnosti a pravidla pro určení členství v domácnosti,
- dotazové instrumenty a způsob práce s nimi,
- časově efektivní vyplňování instrumentů v domácnosti (rozdání papírových dotazníků k vyplnění během osobního rozhovoru s jiným členem domácnosti),
- náročnější či pro další průběh rozhovoru zásadní části dotazníků (např. zjišťování vztahů mezi členy domácnosti a jejich kontrola v dalších vlnách či testy kognitivních schopností ve druhé vlně),
- komunikace s respondenty (vysvětlení způsobu výběru, zajištění anonymity, význam účasti, nenahraditelnost domácnosti ve výběru),
- strategie zvyšování návratnosti, konverze respondentů,
- odměny pro respondenty a tazatele,
- mechanismy kontroly práce tazatelů,
- monitorování terénních prací.

Ve druhém až čtvrtém roce výzkumu se k těmto bodům přidala témata týkající se specifických rysů opakovaného dotazování, zejména:

- vyhledání a kontaktování přestěhovaných domácností a osob,
- vysvětlení významu opakované účasti ve výzkumu,
- evidence nových a bývalých členů domácností,
- podmíněné dotazování v CAPI domácnostním a individuálním dotazníku.

V prvním roce výzkumu se před školením tazatelů uskutečnilo „školení školitelů“, které uspořádali členové výzkumného týmu projektu ze Sociologického ústavu AV ČR, v.v.i. Školení mělo podobu workshopu, při němž pracovníci agentur a výzkumný tým diskutovali a upřesňovali klíčové body pro školení tazatelů.

3.2. Pilotní šetření

V každém ročníku se před hlavním sběrem dat uskutečnilo pilotní šetření výzkumu, jehož cílem bylo otestovat nastavení parametrů výzkumu, především:

- dotazových instrumentů (srozumitelnost, časové nároky, fungování CAPI scriptů),
- motivačních pobídek pro domácnosti a tazatele,
- mechanismů kontroly a monitorování sběru dat v terénu.

V prvním roce výzkumu se pilotní šetření uskutečnilo v období od 24. dubna do 19. května 2015. Dotázáno v něm bylo 302 domácností vybraných postupem obdobným k proceduře využívané pro hlavní sběr.

Pilotní šetření pro druhý rok výzkumu proběhlo od 13. dubna do 9. května 2016. Dotázáno bylo 234 domácností, které se pilotního výzkumu zúčastnily v roce 2015, a 71 nových domácností (celkem tedy 305 domácností). Doplněné domácnosti vybírali tazatelé v okruhu do cca 500 metrů od původních domácností z prvního ročníku, které odmítly zopakovat svoji účast nebo které nebylo možné zastihnout.

Pilotáže třetí vlny se v období od 27. března do 24. dubna 2017 zúčastnilo 272 domácností z těch, které byly dotázány ve druhém roce. Pilotní výzkum ve čtvrtém roce proběhl od 26. března do 24. dubna 2018 na 258 domácnostech ze vzorku vyšetřeného ve třetím roce.

3.3. Avizování výzkumu

V prvním roce výzkumu byly před každou etapou sběru dat s týdenním předstihem rozesílány následující materiály:

- poštou oslovující dopis domácnostem s žádostí o spolupráci a informační leták o výzkumu s kontaktními údaji (telefon, e-mail, web),
- poštou a elektronicky dopis starostům obcí a měst s žádostí o informování obyvatel,
- poštou a elektronicky dopis krajským ředitelstvím Policie ČR informující o aktivitách tazatelů.

Ve druhé až čtvrté vlně obdržely domácnosti předem poštou dopis s žádostí o opětovnou spolupráci včetně informace o výhercích slosování o ceny a leták s vybranými výsledky předchozího ročníku výzkumu a kontaktními údaji.

3.4. Průběh terénních prací

Terénní práce na sběru dat pro **první ročník výzkumu** probíhaly v období od 7. července do 1. prosince 2015. Sběr dat byl rozčleněn do 4 měsíčních etap, v nichž byly tazatelům postupně vydávány adresy domácností. V prvních dvou měsících byla distribuována více než polovina adres (cca 60 %), aby tazatelé mohli opakovat návštěvy v průběhu delšího období a zvýšila se tak šance kontaktu i s hůře dosažitelnými domácnostmi. Kvůli nižší úspěšnosti kontaktu během letních prázdnin (červenec, srpen) byly cca 2 třetiny domácností dotázány od září do prosince 2015.

Z vydaných 13 269 adres bylo dotázáno a kontrolou úspěšně ověřeno 5 159 domácností (42,1 % z adres splňujících požadavky na dotázání, podrobnosti v *tabulce 3*).

Na sběru dat se alespoň jednou dotázanou domácností podílelo 301 tazatelů. V jednotlivých etapách sběru dat bylo tazatelům přidělováno 1 až 6 výběrových bodů. V posledních dvou etapách byly přerozděleny dosavadní výběrové body tazatelů s nízkou aktivitou mezi zbývajícím tazatelem. V průměru na jednoho tazatele připadá cca 15 domácností vyšetřených osobním rozhovorem.

Výběrový soubor 12 750 adres byl rovnoměrně rozdělen mezi agentury MEDIAN a STEM/MARK, které sběr dat provedly se srovnatelnou úspěšností (MEDIAN 2 562 vyšetřených domácností, STEM/MARK 2 597).

Tabulka 3: výsledky kontaktů ve vzorku domácností, vlna 1

vydané adresy	13 269	*)
splňuje požadavky na dotázání („eligible“)	12 265	
dotázáno, ověřeno pomocí kontrol	5 159	42,1 %
odmítnutí	4 216	
dotazování nedokončeno nebo domluveno na termín po skončení terénu	149	
nezastižení (celé domácnosti nebo dospělého člena)	2 265	
domácnost nebyla navštívena	252	
dotázáno, vyřazeno na základě kontrol	224	
nesplňuje požadavky na dotázání („ineligible“)	1004	7,3 %
neschopnost se zúčastnit – žádný člen nemluví česky	14	
neschopnost se zúčastnit – zdravotní stav, stáří	286	
nenalezená adresa	299	
adresa není obydlený byt	405	

*) K výběrovému souboru 12 750 adres bylo během terénních prací doplněno 207 dalších hospodařících domácností na vybrané adrese (více než 1 hospodařící domácnost v bytě) a 312 adres bylo náhodně vybráno jako náhrada téhož počtu adres, které se v terénu ukázaly jako neobydlené či dlouhodobě neobývané.

Ve **druhém roce výzkumu** probíhal sběr dat od 20. června do 31. října 2016. Adresy domácností byly tazatelům vydávány ve 3 fázích trvajících přibližně 6 týdnů. V prvních dvou fázích bylo distribuováno cca 80 % adres. Domácnosti k dotázání byly přidělovány tazatelům postupně podle toho, ve kterých měsících byly vyšetřeny v první vlně. Domácnost nesměla být vyšetřena dříve jak 10 měsíců od účasti v prvním ročníku. Více než polovina domácností (55 %) byla dotázána během června až srpna, třetina domácností v září a zbylých 12 % v říjnu.

K účasti na druhé vlně výzkumu byly přizvány vyšetřené a kontrolami ověřené domácnosti z prvního ročníku (5 093 adres). Počet vydaných adres se v průběhu sběru dat navýšil o 171 domácností, které vznikly rozdělením mateřské domácnosti z prvního ročníku na více samostatných jednotek. To znamená, že u 171 domácností dotázaných ve druhém ročníku bylo zjištěno, že někteří z jejich původních členů odešli a zformovali nové domácnosti. Pouze u 16 z těchto rozdělených domácností se podařilo po dotázání první domácnosti vyšetřit i druhou odvozenou.

Celkem bylo ve druhém roce dotázáno a kontrolou ověřeno 4 147 domácností (79,5 % adres splňujících požadavky pro dotázání, podrobnosti jsou uvedeny v *tabulce 4*). Ze 4 147 zúčastněných domácností bylo 137 (3,3 %) dotázáno na jiné adrese než v prvním roce (změna adresy zahrnuje i přesun do jiného bytu na stejné adrese). Na jinou adresu se přitom mohla přestěhovat jak celá domácnost, tak pouze její vybraní členové. Úspěšnost dotázání domácnosti na jiné adrese činila 47 %.

Sběr dat provádělo 208 tazatelů (alespoň 1 vyšetřená domácnost). Na jednoho tazatele tím v průměru připadá 18 dotázaných domácností. Osobního rozhovoru se v obou letech zúčastnilo 3 727 domácností, z toho 84 % dotázal v obou letech stejný tazatel. Agentury MEDIAN a STEM/MARK se podílely na sběru dat srovnatelnou měrou (MEDIAN 2 050 domácností, STEM/MARK 2 097).

Tabulka 4: výsledky kontaktů ve vzorku domácností, vlna 2

vydané adresy	5 264	
dotázáno ve W1	5 093	*)
vzniklé rozdělením	171	
splňuje požadavky na dotázání („eligible“)	5 217	
dotázáno, ověřeno pomocí kontrol	4 147	79,5 %
odmítnutí	310	
dotazování nedokončeno nebo domluveno na termín po skončení terénu	7	
nezastižení (celé domácnosti nebo dospělého člena)	101	
domácnost nereagovala na CAWI výzvu (jen W1 CAWI domácnosti)	227	
domácnost nebyla navštívena	256	
změna adresy, nedotázáno	163	
dotázáno, vyřazeno na základě kontrol	6	
nesplňuje požadavky na dotázání („ineligible“)	47	0,9 %
neschopnost se zúčastnit – žádný člen nemluví česky	0	
neschopnost se zúčastnit – zdravotní stav, stáří	45	
nenalezená adresa	2	

*) Ve druhé vlně nebylo z 5 159 domácností dotázaných v roce 2016 osloveno 67 domácností, které se zúčastnily prostřednictvím CAWI a nebyl na ně k dispozici funkční e-mailový kontakt. Naopak se do šetření ve druhém roce zapojila i jedna domácnost, která byla kvůli pochybení tazatele vyřazena z dat první vlny.

Sběr dat se ve **třetím roce výzkumu** uskutečnil od 21. června do 30. října 2017. Tazatelům byly postupně dávkovány adresy domácností dle data účasti v předchozím roce tak, aby byl zachován minimálně 10měsíční rozestup mezi dotazováními. Agentura STEM/MARK rozdělila terénní práce do dvou fází, agentura MEDIAN do tří. Během června až srpna byly dotázány téměř dvě třetiny domácností (64 %), během září 29 % a v říjnu zbylých 7 % domácností. V listopadu pokračoval sběr CAWI mutací papírových dotazníků (PAPI dotazníky a časové snímky pro dospělé a děti), které nebyly z důvodu chybného nastavení CAWI rozhraní získány v řádném termínu sběru dat (u části domácností bylo CAWI dotazování uzavřeno ještě před přesměrováním na uvedené dotazníky). Těmto domácnostem byla během listopadu zaslána žádost o dodatečné vyplnění dotazníků. Během listopadu byly takto získány odpovědi v přibližně 20 % CAWI domácností.

K dotázání byly vydány adresy jen těch domácností, které se do výzkumu zapojily ve druhém ročníku a úspěšně prošly kontrolami. Tazatelům bylo k vyšetření přiděleno celkem 4 303 adres, z toho 156 adres vzniklo rozdělením domácností. Obě dceřiné domácnosti byly dotázány v 11 případech.

Ve třetím roce bylo dotázáno a kontrolou ověřeno 3 616 domácností (84,7 % adres splňujících požadavky pro dotázání, podrobnosti obsahuje *tabulka 5*). U 130 (3,6 %) z nich se dotazování uskutečnilo na jiné adrese než ve druhém roce. Úspěšnost dotazování na nové adrese dosáhla 46 %.

Na sběru dat se alespoň jednou vyšetřenou domácností podílelo 170 tazatelů při průměrném počtu přibližně 19 dotázaných domácností na tazatele. Kontinuita tazatele byla zachována u 76 % z 3 195 domácností vyšetřených ve všech třech letech osobním rozhovorem. Dotázané domácnosti byly opět rovnoměrně rozloženy mezi agentury MEDIAN (1794) a STEM/MARK (1822).

Tabulka 5: výsledky kontaktů ve vzorku domácností, vlna 3

vydané adresy	4 303	
dotázáno ve W2	4 147	
vzniklé rozdělení	156	
splňuje požadavky na dotázání („eligible“)	4 269	
dotázáno, ověřeno pomocí kontrol	3 616	84,7 %
odmítnutí	128	
dotazování nedokončeno nebo domluveno na termín po skončení terénu	29	
nezastižení (celé domácnosti nebo dospělého člena)	79	
domácnost nereagovala na CAWI výzvu	108	
domácnost nebyla navštívena	149	
změna adresy, nedotázáno	151	
dotázáno, vyřazeno na základě kontrol	9	
nesplňuje požadavky na dotázání („ineligible“)	34	0,8 %
neschopnost se zúčastnit - žádný člen nemluví česky	0	
neschopnost se zúčastnit - zdravotní stav, stáří	32	
nenalezená adresa	2	

Ve čtvrtém ročníku probíhalo šetření v termínu od 20. června do 15. října 2019. Stejně jako v předchozích letech byly adresy k osobnímu rozhovoru i online dotazování vydávány postupně, aby od účasti ve třetím ročníku uplynulo alespoň 10 měsíců. Agentura STEM/MARK rozdělila terénní práce do dvou fází, agentura MEDIAN do tří. Do konce srpna byla získána data od přibližně 65 % domácností, v září a říjnu bylo došetřeno zbývajících 35 % domácností.

Do vzorku pro dotázání ve čtvrté vlně byly zařazeny všechny domácnosti, které se výzkumu zúčastnily v předchozím roce (3 616). Během terénních prací bylo do vzorku doplněno dalších 118 domácností, které se ustavily oddělením od mateřské domácnosti. Více než jednu dceřinou domácnost se podařilo dotázat v 9 z těchto 118 případů (8 domácností rozděleno na dvě části, jedna na tři).

Ve čtvrtém roce byly získány a kontrolou úspěšně ověřeny údaje od 3 188 domácností a návratnost na úrovni domácností tedy dosáhla 86,4 %. Z těchto domácností se 102 přesunulo od minulého dotazování na novou adresu. Pravděpodobnost vyšetření domácnosti přestěhované na jinou adresu je odhadována na 48 %.

Minimálně jednou dotázanou domácností do výzkumu přispělo 157 tazatelů a jeden tazatel tak v průměru provedl rozhovor v 18 domácnostech. Ve všech čtyřech letech se formou osobního rozhovoru výzkumu zúčastnilo 2 829 domácností. Tentýž tazatel byl ve všech letech zachován v 71 % případů, u ostatních došlo alespoň jednou k výměně tazatele.

Stejně jako v předchozích letech vyšetřily agentury MEDIAN a STEM/MARK srovnatelný počet domácností (1 556 a 1 632).

Tabulka 6: výsledky kontaktů ve vzorku domácností, vlna 4

vydané adresy	3 734	
dotázáno ve W3	3 616	
vzniklé rozdělení	118	
splňuje požadavky na dotázání („eligible“)	3 688	
dotázáno, ověřeno pomocí kontrol	3 188	86,4%
odmítnutí	128	
dotazování nedokončeno nebo domluveno na termín po skončení terénu	5	
nezastižení (celé domácnosti nebo dospělého člena)	29	
domácnost nereagovala na CAWI výzvu	108	
domácnost nebyla navštívena	109	
změna adresy, nedotázáno	111	
dotázáno, vyřazeno na základě kontrol	10	
nesplňuje požadavky na dotázání („ineligible“)	46	1,2%
neschopnost se zúčastnit - žádný člen nemluví česky	0	
neschopnost se zúčastnit - zdravotní stav, stáří	44	
nenalezená adresa	2	

3.5. Kontaktování domácností

V **prvním roce výzkumu** měli tazatelé při pokusu o kontakt s domácností provést minimálně 6 návštěv (rozložených do různých denních dob a částí týdne), z toho 4 v měsíci po přidělení adresy a 2 odložené na pozdější období, aby byla maximalizována šance zastižení domácnosti. První návštěva adresy měla být realizována v týdnu po přidělení výběrového bodu. Systém odměňování tazatelů byl nastaven tak, aby motivoval k uzavření výběrového bodu v měsíci po přidělení.

První kontakt s domácností probíhal formou osobní návštěvy tazatele, telefonická komunikace byla povolena až po navázání prvního kontaktu. Jestliže se tazateli nepodařilo domácnost zastihnout, mohl zanechat ve schránce navštívenku s datem a časem návštěvy a svým kontaktem, aby domácnost měla možnost se s tazatelem sama spojit.

Pro oslovené domácnosti byla v avizním dopise a informačním letáku připravena telefonní čísla a e-mailové adresy pro přímý kontakt s realizátory nebo výzkumným týmem. Na webu výzkumu www.promenyceskespolecnosti.cz byla také vytvořena speciální sekce pro respondenty.

Tazatelé byli vybaveni kopií avizního dopisu a informačního letáku, které v případě potřeby předkládali při návštěvě domácnosti. Všichni tazatelé prošli školením, při němž byla pozornost věnována zásadám komunikace s cílem maximalizovat návratnost (význam výzkumu, způsob výběru, nenahraditelnost domácnosti v šetření, zajištění anonymity, finanční odměny atd.).

Pro účely monitorování terénních prací evidovali tazatelé dny a výsledky jednotlivých kontaktů. Při kontaktu s domácností odmítající účast zaznamenávali tazatelé důvod odmítnutí. Rozlišována byla „tvrdá“ odmítnutí, která znamenala ukončení kontaktu s domácností, a „měkká“ odmítnutí (např. nevhodný čas návštěvy), v jejichž případě se tazatel pokusil domácnost oslovit znovu. Konečný výsledek důvodů odmítnutí v jednotlivých ročnících zobrazuje *tabulka 7*.

Od září 2015 bylo přijato několik dalších opatření na zvyšování šance kontaktu s dosud nezastiženými domácnostmi a konverzi odmítajících domácností. Z veřejných zdrojů byly dohledávány telefonické kontakty na dosud nezastižené domácnosti. Telefonní operátoři poté kontaktovali dané domácnosti a zprostředkovali jejich setkání s tazatelem, pokud domácnost souhlasila s účastí ve výzkumu. Nezastiženým domácnostem byly také znovu poštou zaslány avizní dopisy. V říjnu a listopadu byly odmítající nebo opakovaně nezastižené domácnosti přizvány k účasti k dotazování na internetu (CAWI) pomocí dopisu s unikátním přístupovým kódem do dotazovacího rozhraní.

Ve **druhém až čtvrtém roce výzkumu** se měli tazatelé pokusit zastihnout domácnost minimálně 4krát, z toho 2krát v měsíci po přidělení adresy a 2krát v pozdějším období. K navázání kontaktu jim bylo doporučeno využít telefonické či e-mailové údaje získané od členů domácností v předchozích letech. Ve druhém ročníku motivovalo nastavení tazatelských odměn k oslovení domácností v prvním týdnu po přidělení adresy a uzavření tazatelského úkolu do konce dané fáze sběru dat. Ve třetím a čtvrtém roce nebylo díky vysoké návratnosti podobné opatření potřeba. Jednotlivé pokusy o kontakt, jejich okolnosti a výsledky opět zaznamenávali tazatelé pro účely monitoringu.

Avizní dopis zasílaný ve druhém až čtvrtém ročníku obsahoval telefonický kontakt na konkrétního tazatele, který mohla domácnost využít ke smlouzení schůzky. Stejně jako v prvním roce byli tazatelé vybaveni navštívenkami a kopiemi avizního dopisu a informačního letáku.

V případě přestěhování celé domácnosti nebo některého jejího člena měl tazatel zjistit pomocí informací od ostatních členů domácnosti, sousedů či telefonických a e-mailových kontaktů z první vlny novou adresu či kontaktní údaje přestěhovaných osob a nahlásit přestěhování agentuře. Pokud se nová adresa nenacházela ve vzdálenosti dosažitelné pro původního tazatele, byla přidělena novému tazateli. Za získání kontaktu na přestěhovanou domácnost obdržel tazatel speciální bonus nezávisle na tom, zda domácnost poté sám dotázal. Za úspěšné vyšetření přestěhované domácnosti byl udělován další bonus. Největší překážkou pro dotázání přestěhovaných členů rozdělených domácností byla skutečnost, že ostatní členové původní domácnosti neznali místo pobytu či kontaktní údaje přestěhovaného člena nebo nebyli ochotni tuto informaci poskytnout.

Kromě nabídky přechodu na dotazování po internetu (CAWI konverze) nebyla ve druhém až čtvrtém ročníku využívána žádná speciální opatření pro zvyšování návratnosti.

Tabulka 7: důvody odmítnutí účasti ve výzkumu

	vlna 1	vlna 2	vlna 3	vlna 4
nedostatek času (členové domácnosti stále zaneprázdnění)	10,4 %	13,5 %	28,9 %	35,9 %
špatné načasování, výjimečná situace v domácnosti (právě odjíždějí, nemoc, návštěva atd.)	2,3 %	5,5 %	7,8 %	6,3 %
nezájem / nedůvěra k výzkumům obecně	52,9 %	39,7 %	6,3 %	5,5 %
v nedávné minulosti na nějaký výzkum odpovídali	0,3 %	0,0 %	0,0 %	0,0 %
předchozí špatná zkušenost s výzkumy	1,6 %	1,3 %	0,8 %	0,0 %
nezájem o téma tohoto výzkumu	5,3 %	7,1 %	13,3 %	4,7 %
nedůvěra k tomuto výzkumu (nemá dost informací, vadí mu téma apod.)	3,1 %	1,0 %	1,6 %	0,0 %
obava z nedostatku anonymity / zneužití informací / poskytování osobních údajů	5,9 %	2,6 %	4,7 %	3,9 %
nedostatečná motivace – malá odměna apod.	0,5 %	1,0 %	0,0 %	3,1 %
strach z tazatele (strach otevřít, strach z podomního prodeje)	1,1 %	0,3 %	0,0 %	1,6 %
některý člen domácnosti „zakázal“ účast	4,4 %	8,7 %	6,3 %	3,1 %
jiné (např. domácnost sama uvědomila agenturu, že se nechce zúčastnit)	1,8 %	0,6 %	5,5 %	0,0 %
odmítnutí bez uvedení důvodu	6,8 %	11,9 %	1,6 %	0,0 %
neuvedeno	3,6 %	6,8 %	23,4 %	35,9 %

3.6. CAWI dotazování

V prvním roce výzkumu byla připravena CAWI mutace dotazování jako jedna z konverzních strategií pro odmítající nebo dlouhodobě nekontaktované domácnosti. Dotazníky na webu vyplnilo v prvním ročníku 589 domácností (11 % z celkového počtu dotázaných domácností). Úspěšnost CAWI konverze činila 8 %.

Od druhého ročníku nebyly domácnosti dotázané v předchozím roce pomocí CAWI oslovovány tazatelem k osobnímu rozhovoru, ale byly přizvány přímo ke CAWI dotazování pomocí odkazu v e-mailu či speciálním dopisu. Míra CAWI retence (tj. podíl CAWI domácností z předchozího ročníku, které se v aktuálním roce zúčastnily rovněž pomocí CAWI) činila ve druhém ročníku 49 % a ve třetím i čtvrtém ročníku 70 %. Pouze ve výjimečných případech byly tyto domácnosti na své přání vyšetřeny osobním rozhovorem.

CAWI dotazování bylo od druhého ročníku využíváno jako konverzní strategie u domácností odmítajících osobní rozhovor pouze agenturou MEDIAN. Agentura STEM/MARK ke konverzi z důvodu vysoké retence v osobním dotazování nepřistoupila. Úspěšnost CAWI konverze (tj. podíl domácností, které v aktuálním roce využily nabídku přestoupit z CAPI na CAWI) se pohybovala v rozmezí 25 až 30 %.

Tabulka 8: počty vyšetřených domácností a míry retence a konverze podle způsobu sběru dat

	vlna 1		vlna 2		vlna 3		vlna 4	
CAPI retence	-		3 743		3 222		2 866	
	-		80,7 %		84,9 %		86,8 %	
CAPI konverze	-		1		2		0	
	-		-		-		-	
CAPI celkem	4 570	88,6 %	3 744	90,3 %	3 224	89,2 %	2 866	89,9 %
CAWI retence	-		294		292		284	
	-		48,8 %		69,5 %		70,3 %	
CAWI konverze	589		108		100		38	
	8,3 %		25,7 %		24,9 %		30,9 %	
CAWI celkem	589	11,4 %	402	9,7 %	392	10,8 %	322	10,1 %

3.7. Monitorování terénních prací

Tazatelé zaznamenávali dny, výsledky a další okolnosti návštěv na adresách do kontaktních formulářů a tyto údaje spolu s evidencí vyplněných dotazníků odesílali co nejdříve po provedení návštěv prostřednictvím CAPI dotazového rozhraní. Realizátoři na základě těchto údajů zpracovávali týdenní zprávy o stavu terénních prací, které předávali výzkumnému týmu.

Pracovníci tazatelské sítě pravidelně komunikovali s jednotlivými tazateli o jejich dosavadní a plánované práci na sběru dat. V první vlně výzkumu se z důvodu nižší návratnosti v předchozím období zvýšila intenzita telefonických kontaktů v posledních dvou měsících na minimálně dva za týden.

Kompletní údaje z kontaktních formulářů jsou dostupné v samostatných datových souborech.

3.8. Rozhovor v domácnosti

Doporučovaná strategie pro provedení rozhovorů v domácnostech spočívala v domluvení schůzky na dobu, kdy je v domácnosti přítomno co nejvíce členů, a zajištění co nejvyššího počtu vyplněných instrumentů v rámci jedné návštěvy. Do domácností se však mohli tazatelé opakovaně vracet, aby dokončili rozhovory, dotázali další osoby nebo vyzvedli vyplněné dotazníky na papíře.

Průměrná doba strávená tazatelem v domácnosti činila v první vlně 110 minut, ve druhé až čtvrté vlně 120 minut. V úvodní části návštěvy představili tazatelé výzkum, způsob výběru domácnosti, požadavky na respondenty, náležitosti ochrany osobních údajů a související práva respondentů či systém odměňování za účast. Tazatelé rovněž na úvod sdělili instrukce k vyplnění papírových instrumentů (dotazníků a časových snímků pro dospělé a děti), aby bylo možné dosáhnout efektivního průběhu dotazování. Ve vícečlenných domácnostech vyplňovali za optimálních podmínek ostatní členové své papírové dotazníky, zatímco se tazatel věnoval osobním rozhovorům.

Ke každému dotazníku k samostatnému vyplnění na papíře předal tazatel respondentům propisovací tužku. Pro zajištění co nejladšího průběhu dotazování měl tazatel připravené pastelky a omalovánky, aby jimi pomohl zabavit mladší děti během rozhovoru s dospělými.

V závěru návštěvy sbíral tazatel kontaktní údaje (e-mailové adresy a telefonní čísla) na jednotlivé členy domácnosti pomocí speciálního formuláře s informovaným souhlasem a zajišťoval formální náležitosti potřebné k vyplacení odměny pro domácnost. Tazatel rovněž v závěru kontroloval formální správnost vyplnění papírových dotazníků a časových snímků a případně vyzval respondenty k opravě či doplnění. Pro zúčastněné domácnosti měli tazatelé připravenou kartičku s poděkováním.

3.9. Odměny pro respondenty a tazatele

Domácnosti získaly za účast ve výzkumu finanční odměnu odstupňovanou podle počtu členů domácnosti (od 500 Kč do 1000 Kč). V prvním vlně byl k základní odměně přidáván bonus 100 Kč, pokud byly v domácnosti vyplněny všechny možné dotazové instrumenty. V dalších letech byl po zkušenostech s komplikovanou administrací bonusu a kontrolou správného vyplacení odměn tento bonus zrušen. Svoji odměnu za účast mohla domácnost poukázat na charitativní účely. Před začátkem sběru dat ve druhém až čtvrtém ročníku výzkumu bylo z účastníků předchozí vlny náhodně vybráno vždy 16 domácností, které získaly poukazy na nákup zboží. Ve třetím a čtvrtém ročníku byly za vyplnění dětských dotazníků předávány drobné dárky (mini deskové hry, reflexní samolepky) s cílem posílit návratnost těchto typů instrumentů.

Systém tazatelských odměn byl založen na počtu vyplněných instrumentů v rámci domácnosti. Základní část odměny se vztahovala k vyplnění CAPI dotazníku pro domácnost a prvního CAPI individuálního dotazníku. Za každý další CAPI individuální, případně proxy, dotazník a papírové dotazníky byla vyplácena jednotková sazba stanovená pro každý instrument zvlášť. Tazatelům byly proplaceny jen ty domácnosti, v nichž byla dotázána alespoň polovina členů a vyplněna alespoň polovina všech možných instrumentů.

V jednotlivých letech byl systém tazatelských odměn upravován dle specifických cílů sběru dat. V prvních dvou letech získávali tazatelé bonusové složky odměny za včasné splnění tazatelského úkolu

v přiděleném výběrovém bodě. V prvním roce výzkumu byla od začátku října zvýšena odměna za dotázanou domácnost. Ve druhém až čtvrtém roce výzkumu byly po ukončení sběru dat rozdělovány speciální bonusy nejvýkonnějším tazatelům dle individuálního ohodnocení jejich práce. Ve třetím a čtvrtém roce byla navýšena odměna za získání vyplněných PAPI dotazníků a časových snímků pro děti. Od druhého ročníku výzkumu získávali tazatelé bonus za dodání kontaktních údajů na přestěhované či rozdělené domácnosti.

3.10. Kontrola práce tazatele

K ověření realizace rozhovoru sloužily ve všech vlnách výzkumu dva základní mechanismy:

- náslechy nahrávek vybraných pasáží rozhovoru pořizované se svolením respondenta,
- telefonická, e-mailová nebo korespondenční zpětná kontrola v dotázaných domácnostech, při níž byly ověřovány základní údaje o domácnosti a průběhu rozhovoru.

Těmito kontrolami byly ověřovány domácnosti dotázané tazatelem a u každé z nich se uskutečnil vždy alespoň jeden typ kontroly. Náslechy nahrávek byly v jednotlivých letech provedeny u 85 až 95 % domácností.³ Telefonická, e-mailová či korespondenční zpětná kontrola se v prvním roce odehrála u 94 % domácností. Od druhého ročníku probíhala zpětná kontrola pouze telefonicky, a to u 53 % domácností ve druhé, u 79 % domácností ve třetí a 72 % domácností ve čtvrté vlně. Pokud v rámci jedné kontroly vznikly pochybnosti o správné realizaci dotazování, byl proveden další typ kontroly. Druhou kontrolou prošly také všechny další domácnosti tazatele, u nějž bylo zaznamenáno jakékoli pochybení. Při podezření na provedení rozhovoru s jinou než předepsanou domácností zkontroloval situaci v terénu osobně pracovník tazatelské sítě.

Mimo uvedené dva druhy kontroly realizace rozhovoru probíhala v prvním roce také kontrola navštívení adres (dotázaných i nedotázaných) pomocí unikátních kódů vytištěných na avizních dopisech. Kontrola měla pouze doplňující charakter, protože ne všechny domácnosti byly schopny nebo ochotny předat tazateli doručený dopis.

Do kontrol nebyly zařazeny rozhovory realizované metodou CAWI, protože se uskutečnily na základě unikátních přístupových kódů doručených poštou přímo do vybraných domácností, a to bez zprostředkování tazatelem. Kontrola korektnosti vyplnění CAWI dotazníků je součástí celkové kontroly kvality dat (například konzistence sociodemografických charakteristik mezi vlnami).

3.11. Udržování kontaktu s domácnostmi

Pro respondenty byla připravena webová stránka www.promenyceskespolecnosti.cz, na níž mohli v období sběru dat i mimo něj získat aktuální informace o průběhu výzkumu. Zveřejňovány zde byly jak aktuality o stavu terénních prací, tak drobné analytické výstupy, z nichž některé byly dále mediálně šířeny. Ke sdělování informací a sdílení odkazů na web sloužil profil výzkumu na sociální síti Facebook. Během prosince bylo domácnostem zúčastněným v příslušném ročníku zasláno elektronicky přání k Vánocům společně s informací o úspěšném zakončení sběru dat.

³ Vždy 100 % nahrávek, kde respondent s nahráváním souhlasil a kde nebylo nahrávání znemožněno technickým problémem.

4. Dotazové instrumenty

4.1. Typy instrumentů

Data byla sbírána pomocí standardizovaných dotazníků formou osobního rozhovoru s využitím počítače (CAPI dotazníky) nebo samostatného vyplňování na papíře (PAPI dotazníky a časové snímky).⁴ Pro účely dotazování po internetu byly všechny instrumenty převedeny do webové formy (CAWI).

CAPI dotazník pro domácnost vyplňoval jeden zástupce domácnosti, obvykle osoba nejlépe obeznámená s bytovou a finanční situací domácnosti. Na dotazování se mohl podílet i další člen domácnosti kromě hlavního respondenta, pokud byl o dané otázce lépe informovaný a byl schopný některé údaje opravit, doplnit nebo sám zodpovědět. V dotazníku uváděl respondent jak údaje za celou domácnost, tak údaje o jednotlivých členech domácnosti bez věkového omezení.

CAPI individuální dotazník byl určen všem dospělým členům domácnosti. V případě, že člen domácnosti nebyl schopen dotazník vyplnit (nepřítomnost, zdravotní důvody atp.), jiný dospělý člen domácnosti za něj mohl zodpovědět základní otázky v **CAPI proxy dotazníku**. Podíl proxy dotazníků na celkovém počtu CAPI dotazníků pro dospělé se v jednotlivých vlnách pohyboval kolem 5 %. Průměrné délky CAPI dotazníků v jednotlivých ročnících uvádí *tabulka 9*.

Pro CAPI domácnostní, individuální a proxy dotazování byly připraveny **karty** k vybraným otázkám. Využití karet je značeno v dotaznících u příslušných otázek.

Tabulka 9: průměrná délka CAPI dotazníků v minutách

	vlna 1	vlna 2	vlna 3	vlna 4
CAPI domácnost	16,5	9	11,5	11
CAPI individuální	28	40,5	28,5	31
CAPI proxy	8	9	9,5	7,5

PAPI dotazníky pro dospělé byly předávány všem členům domácnosti starším 18 let, **PAPI dotazníky pro děti** byly určeny dětem od 10 do 17 let. Ačkoli dotazníky pro děti a dospělé obsahují některé společné otázky, jejich obsah se do značné míry liší. V první a druhé vlně byly dětské PAPI dotazníky vytvořeny ve dvou variantách pro věkové skupiny **10–14 let** a **15–17 let**. Varianta pro mladší děti měla jak formátováním, tak počtem zahrnutých otázek vznášet na respondenty menší nároky. Od třetí vlny se z důvodu nízkého počtu dětských respondentů v domácnostech přešlo na jednotný formát analogický dospělému dotazníku. I ve spojeném dětském dotazníku však byla rozlišena základní část pro všechny respondenty a rozšíření pouze pro starší děti.

Formou papírového dotazníku byl administrován i **časový snímek** za předchozí den. Časový snímek měl dvě varianty s různými kategoriemi využívání času, a to **pro děti** od 10 a 17 let a **pro dospělé**.

⁴ Užívání výrazu „PAPI dotazník“ pro dotazník k samostatnému vyplnění na papíře je zdůvodněno v úvodní části dokumentace Shrnutí.

PAPI dotazníky a časové snímky byly rozdány respondentům k samostatnému vyplnění, tazatel pouze vysvětlil způsob práce s nimi. Děti měly své dotazníky zodpovědět bez dohledu rodičů, ale s časovým snímkem mohli rodiče (zvláště mladším) dětem v případě potíží vypomoci.

4.2. Distribuce instrumentů

Při osobním dotazování nebylo určeno pevné pořadí vyplňování instrumentů s výjimkou pravidla, že z CAPI instrumentů je první vyplněn CAPI dotazník pro domácnost. S vyplněním CAPI dotazníku pro domácnost byla v dotazovém rozhraní zároveň založena evidence všech možných instrumentů k vyplnění v dané domácnosti. Tento systém upozorňoval tazatele, kolik dotazníků je možné v domácnosti vyplnit a kolik zbývá, aby byla domácnost považována za úspěšně dotázanou.

PAPI dotazníky a časové snímky mohli respondenti vyplňovat jak před osobním CAPI rozhovorem, tak až po něm. Běžná byla situace, kdy respondenti vyplnili papírové instrumenty během CAPI rozhovoru tazatele s jiným členem domácnosti. V některých případech tazatel předal papírové dotazníky spolu s pokyny k jejich vyplnění při prvním kontaktu s domácností a CAPI rozhovory byly domluveny až na pozdější datum. Docházelo také k situaci, kdy papírové dotazníky vyplnili členové domácnosti až po odchodu tazatele, který je posléze vyzvedl.

V CAWI mutaci dotazování byl jako první zařazen dotazník pro domácnost. Po něm následoval CAPI individuální dotazník, který mohla vyplnit vždy jen jedna osoba v domácnosti. Po ukončení CAPI dotazování byl respondentům nabídnut rozcestník s přístupem k PAPI dotazníkům a časovým diářům pro jednotlivé členy domácnosti. Do těchto dotazníků se vstupovalo jak přímo z rozcestníku, tak z odkazů, které mohly být přeposlány příslušným členům domácnosti na e-mail. Do rozcestníku bylo možné se vrátet.

Protože nebyla stanovena žádná hierarchie vyplňování dotazových instrumentů, může u každého respondenta existovat jakákoli kombinace vyplněných dotazníků, jež byly určeny osobám v jeho věku (s výjimkou kombinace CAPI individuálního dotazníku a CAPI proxy dotazníku). Pouze v domácnostech dotázaných CAWI byla nastavena hierarchie vyplňování popsaná výše.

4.3. Návratnost na úrovni instrumentů

Za úspěšně dotázanou domácnost byla považována taková, v níž bylo dotázáno minimálně 50 % členů domácnosti a vyplněno minimálně 50 % dotazových instrumentů, mezi nimiž musel být CAPI dotazník pro domácnost a alespoň jeden CAPI individuální dotazník. V datech jsou zařazeny také částečně dotázané domácnosti, v nichž byl vyplněn alespoň CAPI dotazník pro domácnost. V prvním ročníku se jedná o 174 domácností, ve druhém o 85, ve třetím 112 a ve čtvrtém o 68 domácností.

Tabulka 10: počty vyplněných instrumentů, vlna 1

CAPI pro domácnost, proměnné na úrovni domácnosti	5 159
CAPI pro domácnost, individuální proměnné	13 083
<i>z toho:</i>	
dospělí	10 476
děti 10 až 17 let	1 079
děti do 10 let	1 528
CAPI individuální a proxy	7 605
<i>z toho:</i>	
individuální	7 118
proxy	487
PAPI dospělí	8 131
PAPI děti	866
<i>z toho:</i>	
15–17 let	306
10–14 let	560
časový snímek dospělí	7 955
časový snímek děti	804

Tabulka 11: míra vyplnění instrumentů celkem a podle typu domácnosti (hhtype), vlna 1

	celkem	domácnost jednotlivce	partnerská/ manželská domácnost	rodič s dospělými dětmi	rodič s dětmi do 18 let	rodiče s dospělými dětmi	rodiče s dětmi do 18 let	ostatní
CAPI individuální bez proxy	67,9 %	99,6 %	71,8 %	63,1 %	85,4 %	52,1 %	67,9 %	50,0 %
CAPI individuální včetně proxy	72,6 %	99,6 %	75,6 %	69,5 %	88,7 %	58,4 %	74,0 %	54,9 %
PAPI dospělí	77,6 %	78,0 %	77,6 %	75,3 %	83,2 %	75,0 %	81,1 %	73,3 %
PAPI děti	80,3 %	-	-	-	79,9 %	-	79,2 %	89,1 %
časový snímek dospělí	75,9 %	77,1 %	76,1 %	72,2 %	81,0 %	74,5 %	79,0 %	70,6 %
časový snímek děti	74,5 %	-	-	-	74,7 %	-	73,6 %	81,8 %

Tabulka 12: počty vyplněných instrumentů, vlna 2

CAPI pro domácnost, proměnné na úrovni domácnosti	4 147
CAPI pro domácnost, individuální proměnné	10 498 ^{*)}
<i>z toho:</i>	
dospělí	8 394
děti 10 až 17 let	886
děti do 10 let	1 218
CAPI individuální a proxy	5 603
<i>z toho:</i>	
individuální	5 270
proxy	333
PAPI dospělí	6 561
PAPI děti	602
<i>z toho:</i>	
15–17 let	209
10–14 let	393
časový snímek dospělí	6 723
časový snímek děti	602

^{*)} Z údajů o počtu členů domácností jsou vyřazeni bývalí členové domácnosti (rstat = 5).

Tabulka 13: míra vyplnění instrumentů celkem a podle typu domácnosti (hhtype), vlna 2

	celkem	domácnost jednotlivce	partnerská/ manželská domácnost	rodič s dospělý- mi dětmi	rodič s dětmi do 18 let	rodiče s dospělý- mi dětmi	rodiče s dětmi do 18 let	ostatní
CAPI individuální bez proxy	62,8 %	99,8 %	66,6 %	55,5 %	84,4 %	47,4 %	61,4 %	43,5 %
CAPI individuální včetně proxy	66,8 %	99,8 %	69,8 %	62,9 %	88,8 %	53,6 %	65,5 %	47,5 %
PAPI dospělí	78,2 %	83,9 %	78,3 %	72,4 %	83,4 %	74,5 %	80,5 %	74,6 %
PAPI děti	67,9 %	-	-	-	61,5 %	-	70,7 %	58,8 %
časový snímek dospělí	80,1 %	82,8 %	81,1 %	75,3 %	84,4 %	77,7 %	81,9 %	76,5 %
časový snímek dětí	67,9 %	-	-	-	62,9 %	-	70,0 %	61,9 %

Tabulka 14: počty vyplněných instrumentů, vlna 3

CAPI pro domácnost, proměnné na úrovni domácnosti	3 616
CAPI pro domácnost, individuální proměnné	9 183 ^{*)}
<i>z toho:</i>	
dospělí	7 357
děti 10 až 17 let	787
děti do 10 let	1 039
CAPI individuální a proxy	4 870
<i>z toho:</i>	
individuální	4 635
proxy	235
PAPI dospělí	5 839
PAPI děti	644
<i>z toho:</i>	
15–17 let	233
10–14 let	411
časový snímek dospělí	5 966
časový snímek děti	625

^{*)} Z údajů o počtu členů domácností jsou vyřazeni bývalí členové domácnosti (rstat = 5).

Tabulka 15: míra vyplnění instrumentů celkem a podle typu domácnosti (hhtype), vlna 3

	celkem	domácnost jednotlivce	partnerská/ manželská domácnost	rodič s dospělý- mi dětmi	rodič s dětmi do 18 let	rodiče s dospělý- mi dětmi	rodiče s dětmi do 18 let	ostatní
CAPI individuální bez proxy	63,0 %	99,9 %	66,3 %	54,3 %	86,0 %	47,9 %	62,2 %	45,2 %
CAPI individuální včetně proxy	66,2 %	99,9 %	69,2 %	59,1 %	88,7 %	52,6 %	65,7 %	48,5 %
PAPI dospělí	79,4 %	88,4 %	77,4 %	75,8 %	86,7 %	78,0 %	80,3 %	76,0 %
PAPI děti	81,8 %	-	-	-	74,5 %	-	83,1 %	82,0 %
časový snímek dospělí	81,1 %	88,3 %	80,5 %	79,2 %	86,0 %	79,6 %	80,8 %	78,6 %
časový snímek děti	79,4 %	-	-	-	68,9 %	-	80,6 %	84,3 %

Tabulka 16: počty vyplněných instrumentů, vlna 4

CAPI pro domácnost, proměnné na úrovni domácnosti	3 188
CAPI pro domácnost, individuální proměnné	8 093 ^{*)}
<i>z toho:</i>	
dospělí	6 532
děti 10 až 17 let	699
děti do 10 let	862
CAPI individuální a proxy	4 222
<i>z toho:</i>	
individuální	4 021
proxy	201
PAPI dospělí	5 132
PAPI děti	572
<i>z toho:</i>	
15–17 let	212
10–14 let	360
časový snímek dospělí	5 391
časový snímek děti	562

^{*)} Z údajů o počtu členů domácností jsou vyřazeni bývalí členové domácnosti (rstat = 5).

Tabulka 17: míra vyplnění instrumentů celkem a podle typu domácnosti (hhtype), vlna 4

	celkem	domácnost jednotlivce	partnerská/ manželská domácnost	rodič s dospělý- mi dětmi	rodič s dětmi do 18 let	rodiče s dospělý- mi dětmi	rodiče s dětmi do 18 let	ostatní
CAPI individuální bez proxy	61,6 %	99,9 %	65,3 %	54,0 %	78,3 %	46,8 %	60,5 %	44,2 %
CAPI individuální včetně proxy	64,6 %	99,9 %	67,6 %	59,9 %	80,6 %	51,8 %	63,3 %	48,1 %
PAPI dospělí	78,6 %	85,5 %	76,3 %	75,7 %	86,0 %	78,1 %	79,0 %	77,5 %
PAPI děti	81,8 %	-	-	-	75,3 %	-	82,2 %	85,9 %
časový snímek dospělí	82,5 %	87,7 %	81,1 %	80,8 %	85,3 %	81,7 %	82,1 %	83,6 %
časový snímek děti	80,4 %	-	-	-	74,1 %	-	81,2 %	81,2 %

Datové soubory zahrnují proměnné uvádějící počty a podíly jednotlivých vyplněných dotazníků v rámci domácnosti (viz oddíl 12. *Proměnné vztahující se k okolnostem dotazování*).

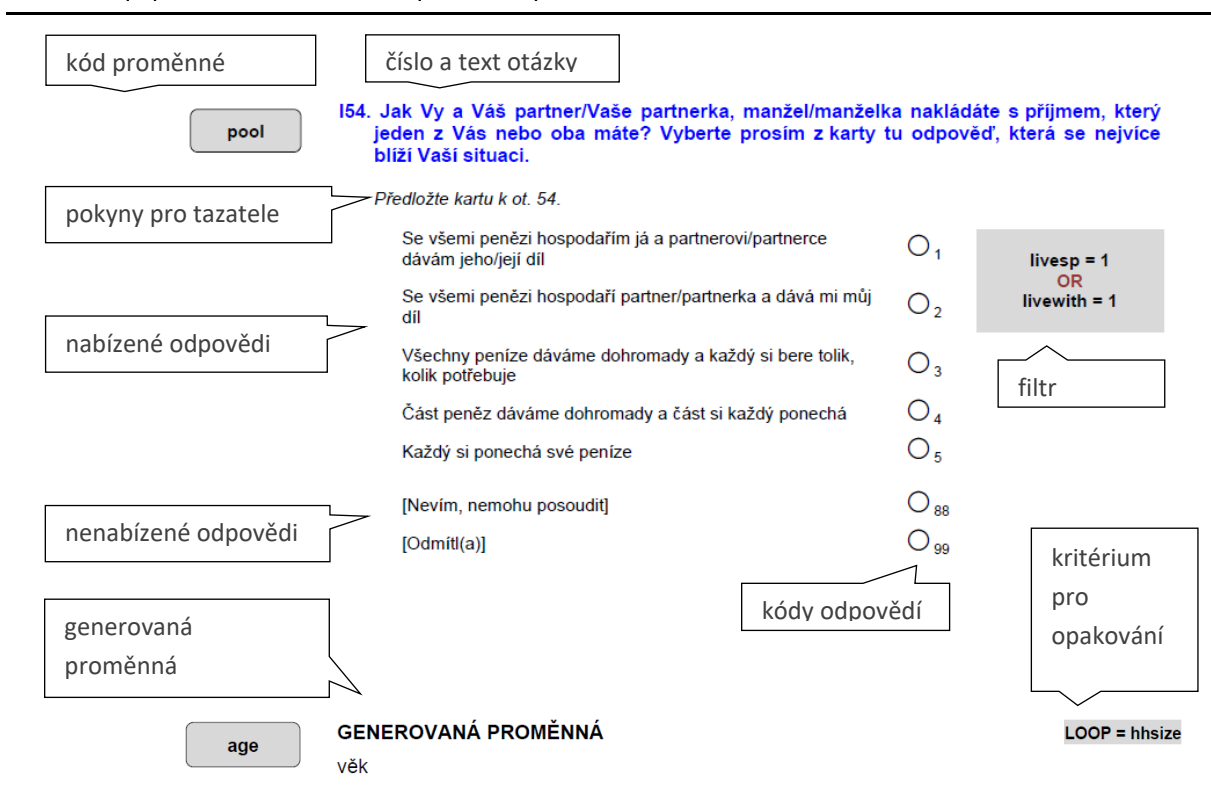
4.4. Dokumentace instrumentů

Labely proměnných v datovém souboru zkráceně vystihují text otázky. Plné znění otázek včetně variant odpovědí je dostupné v jednotlivých dotazových instrumentech. Labely hodnot ve většině případů uvádí plné znění možnosti. Pouze u několika velmi dlouhých popisků byl text zkrácen.

Pro CAPI dotazování byla využívána softwarová řešení agentur MEDIAN, s.r.o., a STEM/MARK, a.s., která umožňovala funkcionálně i vizuálně srovnatelnou implementaci instrumentů. Formální úprava publikovaných CAPI dotazníků se nevztahuje k podobě dotazníků v softwarovém prostředí.

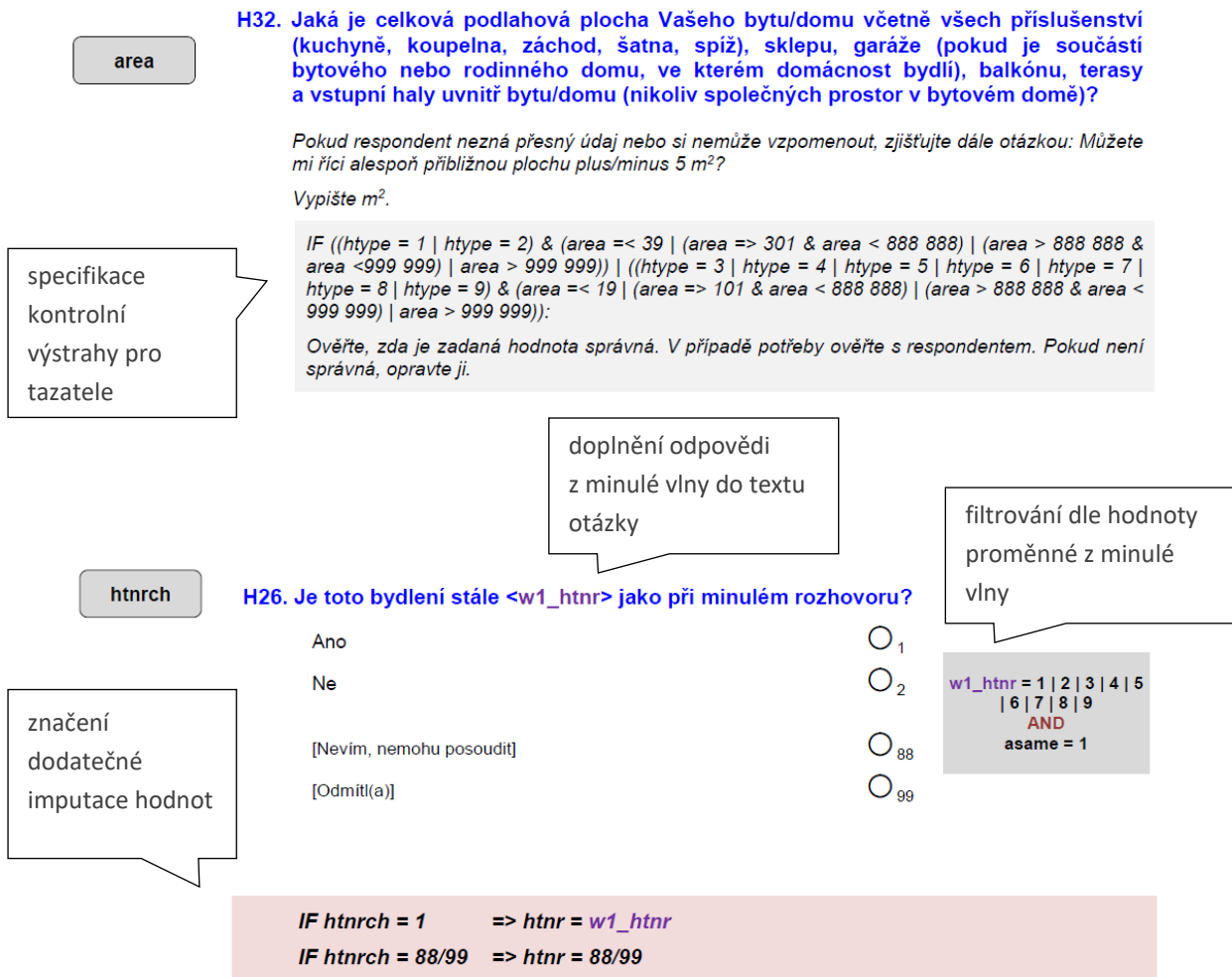
Publikované CAPI dotazníky obsahují kód proměnné, číslo a znění otázky, pokyny pro tazatele, definici filtru, nabízené („platné“) odpovědi a nenabízené („neplatné“) odpovědi. Nenabízené odpovědi jsou uvedeny v hranatých závorkách [...]. V oblasti pro definici filtru je rovněž zahrnuto kritérium pro opakování otázky (LOOP).

Schéma 2: popis základních náležitostí publikovaných CAPI dotazníků



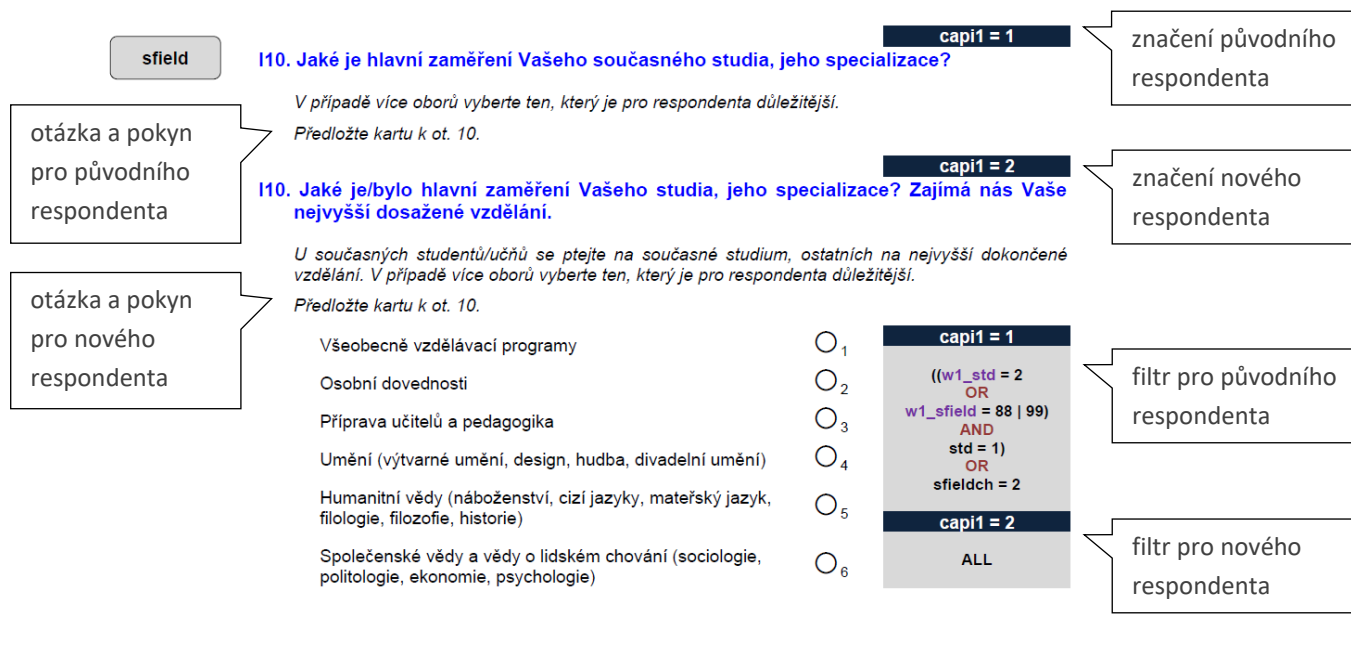
Do dotazníků je dále zahrnuta specifikace kontrolní výstrahy pro tazatele v případě zadání určených nepravděpodobných hodnot a postup dodatečné imputace hodnot do dat u proměnných s využitím podmíněného dotazování.

Schéma 3: další náležitosti publikovaných CAPI dotazníků



V CAPI individuálním dotazníku je od druhé vlny zahrnuto filtrování na základě toho, zda respondent vyplnil tento dotazník v předchozím roce/předchozích letech. Toto filtrování je značeno odděleně od základní filtrační podmínky vždy v záhlaví textu otázky (bílý text v tmavě modrém poli). Pokud se znění otázky liší pro původní a nové respondenty, je otázka uvedena v obou zněních pod příslušnou filtrační specifikací. Stejným způsobem je rozdělena základní filtrační podmínka, pokud není pro staré a nové respondenty stejná. *Schéma 4* ilustruje situaci na proměnné *sfield* z druhého ročníku.

Schéma 4: rozlišení parametrů otázek pro původní a nové respondenty v CAPI INDI dotazníku



PAPI dotazníky jsou publikovány v podobě, v níž byly předloženy respondentům. Doplněny byly pouze kódy proměnných z datového souboru. Doplnění kódů v některých případech vyžadovalo drobné formální úpravy v dotazníku (odsazení začátku textu otázky, v několika případech zmenšení velikosti fontu).

Agentury realizující sběr dat upravily v PAPI dotaznících formát symbolů pro kódování odpovědí s ohledem na svůj postup při pořizování dat. V MEDIANU byly dotazníky skenovány a odpovědi proto musely být označovány křížkem do vyhrazených boxů (ukázka dotazníku ve *schématu 5*). STEM/MARK prováděl ruční pořizování, a v dotazníku této agentury byly proto boxy nahrazeny zvětšeným číselným kódem odpovědi k zakroužkování.

Tištěné časové snímky měly v obou agenturách jednotný formát. Podobně jako u PAPI dotazníků byly do publikovaných verzí doplněny kódy proměnných.

Schéma 5: srovnání původní a publikované podoby PAPI dotazníků

O.1 Když vezmete v úvahu všechny okolnosti, jak jste v současnosti celkově spokojen(a) se svým životem?

naprosto nespokojený(á)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	naprosto spokojený(á)
	0	1	2	3	4	5	6	7	8	9	10	

O.2 Jak moc souhlasíte či nesouhlasíte s následujícími tvrzeními? (zakřížkujte v každém řádku jednu odpověď):

	rozhodně souhlasím	spíše souhlasím	ani souhlas, ani nesouhlas	spíše nesouhlasím	rozhodně nesouhlasím
Často se cítím ve stresu a pod tlakem času.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pravidelně mám dostatek spánku.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Přál(a) bych si mít více času sám(a) pro sebe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Přál(a) bych si mít více času pro rodinu.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Přál(a) bych si mít více času pro rozvoj kariéry.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

lfsat O.1 Když vezmete v úvahu všechny okolnosti, jak jste v současnosti celkově spokojen(a) se svým životem?

naprosto nespokojený(á)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	naprosto spokojený(á)
	0	1	2	3	4	5	6	7	8	9	10	

O.2 Jak moc souhlasíte či nesouhlasíte s následujícími tvrzeními? (zakřížkujte v každém řádku jednu odpověď):

	rozhodně souhlasím	spíše souhlasím	ani souhlas, ani nesouhlas	spíše nesouhlasím	rozhodně nesouhlasím
timestre Často se cítím ve stresu a pod tlakem času.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
timesle Pravidelně mám dostatek spánku.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
timemys Přál(a) bych si mít více času sám(a) pro sebe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
timefam Přál(a) bych si mít více času pro rodinu.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
timecar Přál(a) bych si mít více času pro rozvoj kariéry.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4.5. CAWI mutace dotazových instrumentů

CAWI mutace dotazníků byly připraveny tak, aby co nejlépe odpovídaly původním CAPI a PAPI verzím, avšak specifické rysy dotazování na webu si v několika příkladech vyžádaly úpravu.

Kódy neplatných odpovědí „nevím, nemohu posoudit“ a „odmítl(a)“ v ekvivalentech CAPI dotazníků byly pro respondenty CAWI viditelné, zatímco v osobním rozhovoru tazatel tyto možnosti nenabízel. U otázek *partycl* a *partywv* uváděli respondenti spontánně název politické strany a tazatel pro jejich odpověď vybral odpovídající možnost z nabídky. V CAWI mutaci měl respondent nabídku předloženou. Ve druhém ročníku byla z CAWI mutace CAPI individuálního dotazníku vyloučena převážná většina sekce kognitivních testů.

V ekvivalentech PAPI dotazníků procházeli respondenti na webu jednu otázku po druhé na samostatných obrazovkách (mohli se však k předchozím otázkám vrátet). Časové snímky byly na webu rozloženy do čtyř obrazovek s jednotlivými čtvrtinami dne. V obou typech instrumentů (PAPI dotazníky a časové snímky) museli respondenti zvolit odpověď u všech otázek, zatímco v papírové verzi nemohla být tato kontrola zajištěna.

4.6. Konzistence kódů odpovědí mezi vlnami

Převážná většina otázek opakovaných ve více vlnách výzkumu si zachovává shodnou nabídku možností a jejich číselné kódování. Data z jednotlivých vln obsahují kompletní rozsah labelů využitý napříč vlnami (včetně kódů pro filtry a chybné/chybějící hodnoty), i přestože jednotlivé labely nejsou relevantní pro všechny vlny (např. proměnná *rstat* v první vlně obsahuje label pro kód 5, který byl v dotazníku využit až od druhé vlny).

U několika proměnných bylo mezi vlnami upraveno znění možností, avšak základní význam byl zachován (např. u proměnných *asame* či *rel*). Datové soubory obsahují v popisících hodnot vždy znění možností z příslušné vlny.

U otázek *partycl*, *partyvt* a *partywv* se měnila nabídka politických stran a jejich číslování v dotaznicích využívaných při sběru dat. V datových souborech byly kódy stran harmonizovány a harmonizované kódy jsou uvedeny též u nabídky stran v publikovaných dotaznicích, avšak se zachováním původního řazení stran. V definici filtru proměnné *hclose* obsažené v dotaznicích a také v kartách k dotazníkům jsou ponechány původní číselné kódy. Názvy stran uvedené ve volné odpovědi (*partyclo*, *partywvo* atd.) byly zpětně převedeny do číselného kódu, pokud byla strana přidána do nabídky v následujících letech. V datech se tak objevují i kódy stran, které nebyly v příslušném roce zařazeny mezi uzavřené varianty.

Možnosti otázek *refam*, *rehous* a *reelse* byly pro druhý ročník výzkumu upraveny na základě zkušenosti s kategoriemi využitými v prvním roce. U těchto otázek se liší jak obsah nabízených možností, tak jejich číselné kódování.

U proměnných *psapet* až *psaint* byly ve třetím ročníku upraveny možnosti odpovědí tak, aby se vztahovaly pouze k dění v uplynulém roce, zatímco možnosti nabízené ve druhém ročníku se týkaly událostí v minulosti bez omezení. Kódy odpovědí ve druhém a dalších ročnících tak mají odlišný význam.

Tabulka 18: proměnné s odlišnostmi v kódování odpovědí

asame	1–2	změna znění odpovědi mezi vlnami 2 a 3 se zachováním významu
rstat	1–4	změna znění odpovědi mezi vlnami 1 a 2 se zachováním významu
rstat	5	přidání kódu ve vlně 2
new	1–2	změna znění odpovědi mezi vlnami 2 a 3 se zachováním významu
capistat	1–3	změna kódů generované proměnné mezi vlnami 3 a 4
mstat	2	změna znění odpovědi mezi vlnami 1 a 2 se zachováním významu
educ	0	změna znění odpovědi mezi vlnami 1 a 2 se zachováním významu
rel(x)	3	změna znění odpovědi mezi vlnami 1 a 2 se zachováním významu
rel(x)	5	změna znění odpovědi mezi vlnami 2 a 3 se zachováním významu
rel(x)	8	změna znění odpovědi mezi vlnami 2 a 3 se zachováním významu
rel(x)	10	změna znění odpovědi mezi vlnami 2 a 3 se zachováním významu
rel(x)	18	změna znění odpovědi mezi vlnami 2 a 3 se zachováním významu
dwelob	1–8	změna znění odpovědi mezi vlnami 1 a 2 se zachováním významu
unemre	9	změna uzavřené varianty na vypsání vlastní odpovědi mezi vlnami 1 a 2
partycl	11	změna uzavřené varianty na vypsání vlastní odpovědi mezi vlnami 1 a 2
partycl	12	změna uzavřené varianty na vypsání vlastní odpovědi mezi vlnami 3 a 4
partywv	11	změna uzavřené varianty na vypsání vlastní odpovědi mezi vlnami 1 a 2
partywv	12	změna uzavřené varianty na vypsání vlastní odpovědi mezi vlnami 3 a 4
refam	1–4	změna obsahu možností mezi vlnami 1 a 2
refam	5–10	přidání kódu ve vlně 2
rehous	2	změna znění odpovědi mezi vlnami 1 a 2 se zachováním významu
rehous	4	změna obsahu možnosti mezi vlnami 1 a 2
rehous	5	přidání kódu ve vlně 2
reelse	4	změna obsahu možnosti mezi vlnami 1 a 2
reelse	5–9	přidání kódu ve vlně 2
mdwelob	1–9	změna znění odpovědi mezi vlnami 1 a 2 se zachováním významu
psabou- psaral	1–3	změna obsahu možností mezi vlnami 2 a 3

4.7. Proměnné opakované v cyklech

V CAPI dotazníku pro domácnost a individuálním dotazníku se vybrané otázky opakují pro všechny relevantní osoby, události apod. U těchto proměnných je v dotazníku v oblasti pro filtr uvedeno kritérium opakování otázky (LOOP).

Otázka je opakována tolikrát, kolik činí hodnota kritéria LOOP, a pro ty případy, které vyhovují dalším filtračním podmínkám. Pokud je kritériem pro počet opakování otázek proměnná *hsize*, maximální počet cyklů činí v prvním roce výzkumu 12, ve druhém roce 10. U ostatních číselných proměnných je v dotazníku uveden maximální počet cyklů. Pořadí proměnné v cyklu je značeno příponou (x) v kódu proměnné (po dosažení čísla není již závorka uvedena).

Údaje z cyklovaných otázek jsou v datech z CAPI individuálního dotazníku uloženy v samostatných proměnných, které nesou na konci svého kódu číselné označení cyklu (např. *eduex1*, *eduex2* atd., v dotazníku obecné značení „(x)“). Naopak odpovědi z cyklovaných otázek v CAPI dotazníku pro domácnost jsou uloženy v jedné proměnné u člena domácnosti, kterého se otázka týkala.

Cykly probíhaly ve většině případů v rámci proměnných. Otázka tedy byla zopakována tolikrát, kolik činil relevantní počet případů, a až poté se dotazník přesunul k další otázce. Pouze v případě proměnných zaměstnanecké historie (2. vlna) proběhly vždy všechny otázky týkající se stejného zaměstnání, období nezaměstnanosti atd. a poté dotazník pokračoval etapou dalšího pořadí. V datovém souboru jsou proměnné řazeny tak, jak za sebou otázky při dotazování následovaly.

Speciální případ představují proměnné v úvodní části CAPI dotazníku pro domácnost týkající se charakteristik jednotlivých členů („household grid“). Dotazník měl v této části formu tabulky a tazatel/respondent mohl vyplňovat údaje po osobách i po proměnných.

4.8. Podmíněné dotazování

Ve druhé až čtvrté vlně bylo v CAPI domácnostním a individuálním dotazníku využito tzv. podmíněné dotazování („dependent interviewing“). Průchod dotazníkem či znění otázek se tedy odvíjely od informací získaných v minulých letech.

Řada otázek byla v obou instrumentech filtrována na základě hodnot proměnných z předchozích vln. CAPI dotazník pro domácnost začínal vždy zobrazením adresy z minulé vlny a výzvou k doplnění informace, zda domácnost žije stále na stejné adrese jako v minulém roce. Tato proměnná *asame* pak hrála klíčovou roli v průchodu domácnostním dotazníkem. Pokud domácnost zůstala na stejné adrese, nebyly pokládány otázky na stabilní charakteristiky bydlení. Průchod individuálním CAPI dotazníkem zásadně určovala skutečnost, zda osoba odpovídala na CAPI individuální dotazník v alespoň jedné z předchozích vln. Pokud ano (tzv. původní respondent), nedostávala znovu otázky na rodinu původu či vzdělanostní historii a na tyto okolnosti byli dotázáni pouze noví respondenti. U nových respondentů nebyla naopak ve druhé vlně zařazena kompletní zaměstnanecká historie, aby nedocházelo k přílišnému prodlužování délky dotazování.

Otázky na klíčové údaje o osobách a domácnostech (přehled členů domácnosti a jejich základní sociodemografické proměnné, právní důvod užívání bydlení, pracovní postavení atp.) nebyly pokládány

znovu, ale respondenti potvrzovali platnost údajů z posledního dotazování nebo jim byl údaj před vlastní odpovědí připomenut (tzv. proaktivní podmíněné dotazování). Reaktivní podmíněné dotazování, při němž je předchozí hodnota respondentovi připomenuta, pokud se odchyluje od nové odpovědi, využito nebylo.⁵

Ve druhém ročníku byly při podmíněném dotazování využity hodnoty proměnných z první vlny. Příslušné proměnné jsou ve filtrech či textech otázek v dotaznících z druhé vlny značeny předponou w1 před podtržítkem (např. *w1_name*) a fialovou barvou fontu (viz *schéma 3*). Pro podmíněné dotazování ve třetím a čtvrtém ročníku byly zkonstruovány pomocné proměnné kombinující informace získané v předchozích vlnách. Tyto pomocné proměnné jsou v dotaznících ze třetí a čtvrté vlny značeny předponou o před podtržítkem (např. *o_name*) a fialovou barvou fontu. Vložení odpovědi z minulé vlny do textu otázky je značeno uvedením kódu proměnné ve špičatých závorkách, např. *<w1_htr>* (viz *schéma 3*). Modifikace znění otázky podle hodnoty proměnné z předchozího ročníku je v textu příslušné otázky značena červeným fontem.

4.9. Kognitivní testy (vlna 2)

Ve druhé vlně výzkumu bylo do dotazování zahrnuto několik úkolů prověřujících kognitivní schopnosti respondentů. Kognitivní testy byly součástí CAPI individuálního dotazníku a byly tedy předloženy jen respondentům toho instrumentu. Protože testy nejsou převoditelné do podoby nevyžadující přítomnost tazatele, nebyly zařazeny do CAWI mutace dotazování. CAWI respondenti odpovídali jen na úvodní otázku na celkové zhodnocení paměti (proměnná *selfmem*) a v ostatních proměnných ze sekce kognitivního testování mají přiřazen kód 97 / 999 997.

Podoba testů byla přejata ze třetí vlny UK Household Longitudinal Study (Understanding society).⁶ Texty otázek a pokynů pro tazatele byly co nejvíce zachovány. Nejvýraznější změnami prošlo zadání úkolů prověřujících numerické schopnosti, v nichž byly číselné údaje upraveny tak, aby odpovídaly českému kontextu. Zároveň nebyl průchod těmito úlohami určován úspěšností v předchozí numerické úloze.

Kognitivním testům byla věnována samostatná část školení tazatelů. Tazatelé se měli vyvarovat prezentování úkolů jako testů inteligence a naopak je měli podat jako zpestřující složku dotazování. Tazatelé byli striktně instruováni neukazovat respondentům v této části dotazníku obrazovku a jakkoli jim radit a pomáhat. Ačkoli snahou tazatelů bylo zajistit respondentům pro vyplňování testů soukromí bez přítomnosti jiných členů domácnosti, nebylo toho vždy možné dosáhnout.

V testu epizodické paměti respondent nejprve vyslechl nahrávku s 10 slovy a po jejím skončení měl v libovolném pořadí zopakovat co nejvíce slov, která si zapamatoval. Čas na vybavení slov nebyl omezený. V případě technických problémů se spuštěním nahrávky četl slova tazatel tempem srovnatelným s tempem nahrávky. Test byl připraven ve dvou variantách s odlišnými desítkami slov.

⁵ O typech podmíněného dotazování pojednává např. 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.

⁶ McFall, S. 2013. *Understanding Society: UK Household Longitudinal Study: Cognitive Ability Measures*. Colchester: University of Essex, Institute for Social and Economic Research.

Každá z variant byla náhodně položena polovině respondentů. Sadu slov využitou u daného respondenta určuje proměnná *wordver*.

V testu slovní plynulosti spustil tazatel po přečtení pokynů časomíru odpočítávající 60 vteřin, během nichž měl respondent vyjmenovat co nejvíce názvů zvířat. Tazatel poté zaznamenal do dotazníku počet zvířat vyjmenovaných v daném limitu. Během limitu zapisoval tazatel jednotlivé názvy zvířat (případně zkratky či čárky, pokud nestihl zapsat celá slova) na zvláštní arch s instrukcemi pro správné zadání a vyhodnocení úlohy. Pro účely kontroly správného vyhodnocení počtu vyjmenovaných zvířat byla tato úloha také nahrávána. Údaj zadaný tazatelem byl při čištění dat ve výzkumných agenturách porovnáván se záznamovými archy a nahrávkami a v případě potřeby opraven.

V úlohách prověřujících numerické schopnosti řešili respondenti 4 příklady zaměřené na praktickou matematickou gramotnost. Poté, co tazatel přečetl zadání úlohy, měl respondent samostatně zformulovat odpověď. Tazatel nakonec označil, zda se respondentova odpověď shoduje se správným řešením uvedeným u otázky („správná odpověď“) či nikoli („špatná odpověď“).

V kognitivních úlohách (s výjimkou vlastního zhodnocení paměti, *selfmem*) neměl tazatel k dispozici kódy neplatných odpovědí „nevím, nemohu posoudit“ a „odmítl(a)“.

4.10. Zaměstnanecká historie (vlna 2)

CAPI individuální dotazník zahrnoval ve druhé vlně sekci mapující kompletní historii respondentů na trhu práce. Otázky však byly položeny jen těm respondentům, kteří CAPI individuální dotazník vyplnili již v prvním roce. Protože noví respondenti CAPI individuálního dotazníku zodpovídali otázky týkající se vzdělanostní historie či rodinu původu, které byly pro původní respondenty vypuštěny, jejich dotazník byl časově náročnější.

Při popisu zaměstnanecké historie nepostupovali respondenti chronologicky, ale místo toho nejprve uvedli počty jednotlivých období (např. počet zaměstnání trvajících déle než 1 rok) a posléze se v dalších otázkách vraceli k podrobnějším charakteristikám jednotlivých období.

Aby byla respondentům usnadněna orientace v otázkách a podpořeno vybavování jednotlivých období a přesnější časové zařazení událostí, mohli tazatelé využít pomocnou osu k záznamu zaměstnanecké historie. Na první straně dokumentu byla vytištěna časová osa sahající od roku 1955 po rok 2016 se samostatnými oddíly pro záznam: a) základní vojenské služby, b) zaměstnání, c) období nezaměstnanosti, d) mateřských/rodičovských dovolených, a polí pro zápis počtu těchto období. Druhá strana materiálu obsahovala příkladové popisy dvou různých zaměstnaneckých historií a odpovídající vyplnění časové osy.

Využití časové osy jako podkladu pro vyplňování otázek bylo dobrovolné a na rozdíl od archu pro zápis názvů zvířat, PAPI dotazníků a časových snímků je tazatelé nazasílali zpět do výzkumných agentur.

5. Obsah a struktura datového souboru

Datové soubory, k nimž se vztahuje tato dokumentace, obsahují proměnné z CAPI a PAPI dotazníků pro dospělé i děti. Proměnné z časových snímků jsou uloženy v samostatných datových souborech s vlastní doplňující dokumentací. Kompletní přehled proměnných v datech je dostupný v samostatném MS Excel souboru.

5.1. Řazení proměnných

Proměnné z jednotlivých instrumentů za sebou následují podle pořadí v *tabulce 19*. Proměnné, které byly dotazovány ve více instrumentech, jsou v datovém souboru zařazeny mezi proměnné z dotazníku dřívějšího pořadí. Výjimkou jsou proměnné *health* a *hamper*, dotazované v CAPI proxy dotazníku a v PAPI dotazníku pro dospělé, které jsou umístěny mezi proměnnými z PAPI dotazníku pro dospělé. Pokud osoba vyplnila PAPI dotazník pro dospělé a zároveň byl za ni zodpovězen CAPI proxy dotazník, byla u proměnných *health* a *hamper* do dat vložena hodnota z PAPI dotazníku pro dospělé (s výjimkou případu, kdy otázka zůstala v PAPI dotazníku bez odpovědi).

Labely proměnných obsahují písmeno značící dotazník a číslo odpovídající otázky z daného instrumentu. Identifikační čísla, proměnné zkonstruované z hodnot jiných proměnných (generované) a proměnné vztahující se k průběhu dotazování toto značení nemají.

Tabulka 19: pořadí skupin proměnných v datovém souboru

identifikace domácnosti/jednotlivce	
CAPI dotazník pro domácnost	H + číslo otázky v labelu proměnné
CAPI individuální a proxy dotazník	I, P + číslo otázky v labelu proměnné
PAPI dotazník pro dospělé	A + číslo otázky v labelu proměnné
PAPI dotazník pro děti 15–17 let	Y + číslo otázky v labelu proměnné (od vlny 3 YC)
PAPI dotazník pro děti 10–14 let	C + číslo otázky v labelu proměnné (od vlny 3 YC)
stratifikační postavení	
místo bydliště domácnosti	
charakteristiky dotazování	
časy dotazování	
vyplnění dotazníků v domácnosti	
indikace vyplnění instrumentů a váhy	

5.2. Struktura souboru

Případy v datových souborech reprezentují členy dotázaných domácností (13 077 v první vlně, 10 773 ve druhé vlně atd.). Pro každého člena domácnosti jsou dostupné domácnostní a individuální údaje z CAPI dotazníku pro domácnost. Hodnoty proměnných na úrovni domácnosti jsou všem členům domácnosti společné. Výjimku představují osoby, které v době dotazování už nebyly členy dané domácnosti. U nich se u většiny proměnných na úrovni domácností objevuje neplatný kód 97 / 999 97 (podrobnosti k bývalým členům uvádí oddíl 7. *Proměnné popisující změnu složení domácnosti*).

Hodnoty proměnných z dalších dotazníků (CAPI individuální a proxy dotazník, PAPI dotazníky) jsou dostupné jen u respondentů, kteří vyplnili daný dotazový instrument, v opačném případě je hodnota prázdná (system missing). Strukturu datového souboru ilustruje *tabulka 20*.

Tabulka 20: struktura datového souboru

hid	pid	CAPI pro domácnost, úroveň domácnosti	CAPI pro domácnost, individuální úroveň	CAPI individuální/proxy	PAPI dospělí	PAPI děti
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. Proměnné nezařazené do datového souboru

V datových souborech nejsou zařazeny všechny otázky/proměnné z dotazových instrumentů. Vynechány byly proměnné poskytující detailnější informace o respondentovi (např. křestní jméno, které při dotazování sloužilo k rozlišování členů domácnosti; popis pracovní činnosti, adresa předchozího bydliště) a údaje sloužící ke kontrole v průběhu dotazování (např. *namech*). Tyto proměnné nezařazené do datového souboru shrnuje *tabulka 21*.

Dále byly vynechány proměnné vztahující se k 11. a 12. členům domácnosti (vyšetřené domácnosti měly maximální 10 členů, od druhého roku výzkumu už dotazníky umožňovaly záznam nejvíce 10 členů domácnosti) a proměnné partnerské historie z PAPI dotazníku pro dospělé ve 2. vlně vztahující se k partnerům nevyužitého pořadí (7 až 10).

Tabulka 21: proměnné nezařazené do datového souboru			W1	W2	W3	W4
achange	objasnění změny adresy	CAPI dotazník pro domácnost	-	-	1	1
name	křestní jméno (případně přezdívka)	CAPI dotazník pro domácnost	1	1	1	1
namech	kontrola zadání správného jména	CAPI individuální dotazník	1	1	1	1
sexch	kontrola zadání správného pohlaví	CAPI individuální dotazník	1	1	1	1
birthmch	kontrola zadání správného měsíce narození	CAPI individuální dotazník	1	1	1	1
birthych	kontrola zadání správného roku narození	CAPI individuální dotazník	1	1	1	1
educch	kontrola zadání správného vzdělání	CAPI individuální dotazník	1	1	1	1
stdych	kontrola zadání správného studijního stavu	CAPI individuální dotazník	1	1	1	1
estatch	kontrola zadání správného ekonomického postavení	CAPI individuální dotazník	1	1	1	1
mstatch	kontrola zadání správného rodinného stavu	CAPI individuální dotazník	1	1	1	1
sname(x)	název školy studované v současnosti	CAPI individuální a proxy dotazník	1	1	1	1
sstr(x)	ulice školy studované v současnosti	CAPI individuální a proxy dotazník	1	1	1	1
smun(x)	obec školy studované v současnosti	CAPI individuální a proxy dotazník	1	1	1	1
sprog(x)	obor studovaný v současnosti	CAPI individuální a proxy dotazník	1	1	1	1
swrkps	popis pracovní činnosti v období mezi vlnami	CAPI individuální dotazník	-	1	1	1
nwrkps	popis pracovní činnosti budoucí práce	CAPI individuální dotazník	-	1	1	1
wrkps	popis pracovní činnosti respondenta	CAPI individuální a proxy dotazník	1	1	1	1
mstatcor	zda rodinný stav stejný jako při minulém dotazování	CAPI individuální dotazník	-	1	1	1
mcorch	rodinný stav při minulém dotazování	CAPI individuální dotazník	-	1	1	1
pwrkps	popis pracovní činnosti partnera respondenta	CAPI individuální dotazník	-	1	1	1
liwhy	jiný důvod, proč respondent nežil do 16 let převážně s oběma biologickými rodiči	CAPI individuální dotazník	1	1	1	1
fwrkps	popis pracovní činnosti otce respondenta v době, kdy respondentovi bylo 16 let	CAPI individuální a proxy dotazník	1	1	1	1
mwrkps	popis pracovní činnosti matky respondenta v době, kdy respondentovi bylo 16 let	CAPI individuální a proxy dotazník	1	1	1	1

mstreet	ulice předchozího bydliště	CAPI individuální dotazník	1	1	1	1
mcounty	okres předchozího bydliště	CAPI individuální dotazník	1	1	1	1
mmunic	obec předchozího bydliště	CAPI individuální dotazník	1	1	1	1
mpsc	PSČ předchozího bydliště	CAPI individuální dotazník	1	1	1	1
hwrkps(x)	popis pracovní činnosti v minulosti	CAPI individuální dotazník	-	1	-	-
papiname	křestní jméno vyplněné v PAPI dotazníku	PAPI dospělí a děti	1	1	1	1
papiage	věk vyplněný v PAPI dotazníku	PAPI dospělí a děti	1	1	1	1
papisex	pohlaví vyplněné v PAPI dotazníku	PAPI dospělí a děti	1	1	1	1
jobas	popis práce, kterou chce dítě v budoucnu vykonávat	PAPI děti	1	1	1	1
jobex	popis práce, kterou si dítě myslí, že bude v budoucnu vykonávat	PAPI děti	1	1	1	1

6. Identifikační proměnné

Domácnostem a osobám jsou přiřazena čísla umožňující unikátní identifikaci jednotlivých případů. Identifikační čísla domácností a jednotlivců jsou specifická pro danou vlnu a při napojování souborů z více vln je nutné používat vždy jen odpovídající identifikátory. Příslušnost identifikátoru k vlně značí jeho předpona (např. *w1_hid*). Pro spojení osob z první a druhé vlny je tedy třeba využít proměnnou *w1_pid* z první vlny ve spojení s proměnnou *w1_pid* z druhé vlny (nikoli proměnnou *w2_pid* z druhé vlny) a obdobně postupovat při spojování dat z dalších vln.

Členům domácností jsou v jednotlivých datových souborech přiřazeny identifikátory ze všech předchozích, ale nikoli následujících vln. Hodnoty identifikátorů jsou však unikátní právě jen v dané vlně. Např. v datovém souboru z druhého roku výzkumu se v identifikátoru *w1_pid* vyskytují duplicitní hodnoty, které vznikají v důsledku vyšetření více než jedné dceřiné domácnosti u rozdělených domácností. Vznikem duplicit a rozlišením primárního a sekundárního výskytu osoby v rámci dané vlny se blíže zabývá oddíl 7. *Proměnné popisující změnu složení domácnosti*.

Domácnosti jsou odlišeny unikátními identifikačními čísly (*hid*), která jsou společná všem členům jedné domácnosti. Ve druhém ročníku výzkumu je v identifikaci *hid* z první vlny (*w1_hid*) nahrazena číslice 0 na druhé pozici hodnotou proměnné *asame*. Pokud se původní domácnost rozdělila a všechny její části zůstaly na stejné adrese, objevuje se na druhé pozici proměnné *hid* arbitrárně stanovené pořadí domácnosti. Stejným způsobem jsou zkonstruovány identifikátory domácností ve třetí a čtvrté vlně. Tedy *w3_hid* vzniká na základě *w2_hid*, v němž je číslice 0 na třetí pozici nahrazena hodnotou *asame* ze třetí vlny a *w4_hid* je tímto způsobem odvozen z *w3_hid* a *asame* ze čtvrté vlny. Pokud byla hodnota *asame* z předchozí vlny zpětně opravována, do příslušného identifikátoru *hid* už zasahováno nebylo. V několika případech se tak údaj *asame* a číslice v identifikátoru neshodují.

Unikátní identifikační číslo osoby (*pid*) vzniká v každé vlně jako kombinace identifikace domácnosti (*hid*) a pořadí osoby v rámci dané domácnosti (*pno*). Unikátní identifikační číslo osoby (*pid*) určuje respondenta CAPI individuálního a PAPI dotazníků a identifikuje osobu, za niž byl vyplněn CAPI proxy dotazník. Identifikační číslo respondenta CAPI proxy dotazníku (tedy osoby, která na dotazník odpovídala, nikoli osoby, k níž se vztahují odpovědi) obsahuje proměnná *pidpro*, identifikační číslo respondenta CAPI dotazníku pro domácnost obsahuje proměnná *pidhr*.

Tabulka 22: identifikační proměnné jednotlivce/domácnosti

w(x)_hid	unikátní identifikační číslo domácnosti ve vlně (x)
w(x)_pid	unikátní identifikační číslo osoby ve vlně (x)
pno	pořadí osoby v domácnosti
pidhr	identifikační číslo osoby, která vyplnila dotazník pro domácnost
pidpro	identifikační číslo osoby, která vyplnila proxy dotazník

K jednotlivým osobám jsou v datech přiřazena identifikační čísla dalších členů domácnosti, kteří mají k osobě vybraný vztah (*tabulka 23*).

Tabulka 23: identifikační čísla dalších členů domácnosti

pidp	identifikační číslo člena domácnosti, který je manžel/manželka (<i>rel</i> = 1), stálý partner/stálá partnerka (<i>rel</i> = 2), registrovaný partner/registrovaná partnerka (<i>rel</i> = 3) osoby
pidch(x)	identifikační číslo člena domácnosti, jehož <i>pno</i> je (x) a který je biologické (<i>rel</i> = 4), nevlastní (<i>rel</i> = 5) nebo adoptivní (<i>rel</i> = 6) dítě osoby nebo jeho dítě v pěstounské péči (<i>rel</i> = 7)
pidf *)	identifikační číslo muže (<i>sex</i> = 1) v domácnosti, který je biologický (<i>rel</i> = 9), nevlastní (<i>rel</i> = 10) nebo adoptivní (<i>rel</i> = 11) otec nebo pěstoun (<i>rel</i> = 12) osoby
pidm *)	identifikační číslo ženy (<i>sex</i> = 2) v domácnosti, která je biologická (<i>rel</i> = 9), nevlastní (<i>rel</i> = 10) nebo adoptivní (<i>rel</i> = 11) matka nebo pěstounka (<i>rel</i> = 12) osoby
pidgf	identifikační číslo muže (<i>sex</i> = 1) v domácnosti, který je dědeček (<i>rel</i> = 21) osoby
pidgm	identifikační číslo ženy (<i>sex</i> = 2) v domácnosti, která je babička (<i>rel</i> = 21) osoby

*) Osobám, které mají v domácnosti nevlastního rodiče stejného pohlaví jako pohlaví biologického rodiče, je přiřazeno identifikační číslo také v proměnných vázících se k druhému pohlaví, než je pohlaví biologického rodiče (např. v proměnné *pidf* se nachází identifikační číslo partnerky biologické matky).

7. Proměnné popisující změnu složení domácnosti

Změnu ve složení domácnosti mezi po sobě následujícími vlnami šetření odráží několik proměnných, které byly položeny jako otázky nebo generovány z hodnot jiných proměnných (*tabulka 24*). Tyto proměnné jsou zahrnuty v datových souborech z druhého a dalších ročníků (kromě proměnné *rstat* přítomné již v první vlně).

Ve druhé až čtvrté vlně se v úvodní části CAPI dotazníku zobrazila tabulka členů domácnosti („household grid“) z minulého roku (křestní jména, pohlaví, věk atd.). V proměnné *rstat* (status člena domácnosti) poté měli tazatelé a respondenti zaznačit, zda daná osoba stále k domácnosti patří (kategorie 1 až 4) či z různých důvodů již jejím členem není (úmrtí, přestěhování atd.). Osoby s hodnotou 5 (Už není členem mé domácnosti) v proměnné *rstat* jsou označovány jako bývalí členové domácnosti. Ve třetí a čtvrté vlně byli do tabulky v dotazníku zahrnováni i bývalí členové z loňských let s jedinou výjimkou, a to pokud zemřeli (*leftre* = 1). U bývalých členů z minulých let mohlo být pomocí aktuální proměnné *rstat* uvedeno, že opět k domácnosti patří (tzv. navrátní se bývalí členové, „rejoiners“). Do tabulky tazatelé a respondenti dopisovali všechny nové členy, kteří dosud v žádném roce k domácnosti nepatřili.

Proměnná *hmem(x)* určuje, zda byla daná osoba členem domácnosti v ročníku, který odpovídá číselnému indexu na konci proměnné. Proměnná není nikdy dostupná pro současný rok (např. *hmem2* v souboru z druhé vlny), protože v aktuální vlně určuje status respondenta proměnná *rstat*. Hodnotu 1 (Ano) mají v proměnné *hmem(x)* osoby, které byly v dané vlně označeny jako rezidenti nebo dočasně nepřítomní členové domácnosti (*rstat* = 1/2/3/4/88/99). Hodnotou 2 (Ne) jsou v příslušné proměnné *hmem(x)* značeny dva typy osob. Jednak jsou to bývalí členové, tedy osoby, které před dotazováním v dané vlně domácnost opustily (*rstat* = 5). Do kategorie 2 v proměnných *hmem(x)* zadruhé spadají osoby, které se členy domácnosti staly až v budoucích vlnách (narození, přistěhování atd.), tedy v současné vlně zatím k domácnosti nepatří. Tyto osoby mohou být také rozpoznány pomocí kódu 97 v individuálním identifikátoru *pid* z příslušné vlny.

Do datových souborů jsou vždy kromě současných členů domácnosti zahrnuti i všichni bývalí členové (včetně zemřelých). V důsledku této skutečnosti dochází při vyšetření obou částí rozdělených domácností k duplicitám. Členové těchto domácností jsou v datech obsaženi dvakrát, a to jednou jako bývalí členové svých původních domácností a jednou jako současní členové svých nových domácností. Identifikace osoby (*pid*) ze současného ročníku je vždy unikátní (díky odvození z identifikace domácnosti *hid*), avšak identifikátory z předchozích let nikoli. Pro rozlišení „primárního výskytu“ (tj. v domácnosti, jejímž je osoba aktuálním členem) slouží proměnná *primary*. Proměnná *primary* se vztahuje vždy jen k aktuální vlně. Případ se sekundárním výskytem v dané vlně může být v následující vlně opět primární.

Přítomnost alespoň jednoho nového člena v domácnosti ve srovnání s minulým ročníkem indikuje proměnná *new*. Návazná proměnná *newn* pak určuje počet nových členů. Od třetího ročníku jsou mezi nové členy zařazeny také osoby, které domácnost v průběhu výzkumu opustily a poté se do nich vrátily (navrátní se bývalí členové). Obdobně proměnná *leavers* vyjadřuje počet bývalých členů domácnosti. Ve třetím a čtvrtém ročníku výzkumu jsou mezi bývalé členy ovšem započítáni pouze ti, kteří domácnost opustili mezi bezprostředně předcházející a současnou vlnou. Proměnné *new*, *newn* a *leavers* jsou na rozdíl od předchozích individuálních proměnných vyjádřeny na úrovni domácnosti.

Proměnná na úrovni domácnosti *hchange* značí, zda se změnilo či nezměnilo složení domácnosti. Pokud se složení domácnosti změnilo, znamená to, že do domácnosti přibyl alespoň jeden nový člen nebo alespoň jeden domácnost opustil.

Tabulka 24: proměnné vztahující se k novým a bývalým členům domácnosti

rstat	ke kategoriím 1 až 4 využívaným v prvním ročníku pro určení residenčního statusu člena domácnosti byla od druhého ročníku doplněna kategorie 5 identifikující bývalého člena domácnosti
hhmem(x)	ukazatel členství osoby v domácnosti ve vlně (x)
primary	výskyt osoby ve vzorku: primární (člen nerozdělených domácností, aktuální člen rozdělené domácnosti) nebo sekundární (bývalý člen rozdělené domácnosti). Hodnota vypovídá pouze o aktuální vlně (je tedy možný přechod ze sekundárního výskytu opět do primárního).
new	přítomnost nových členů v domácnosti (včetně navrátilivších se bývalých)
newn	počet nových členů v domácnosti (včetně navrátilivších se bývalých)
leavers	počet bývalých členů domácnosti (domácnost opustili mezi aktuální a bezprostředně předcházející vlnou)
hchange	indikace změny ve složení domácnosti (tj. přítomnost nových nebo bývalých členů domácnosti)

Bývalým členům domácnosti je v proměnných z CAPI dotazníku pro domácnost přidělena platná hodnota (při splnění ostatních filtračních podmínek) v proměnných *w(x)_hid*, *w(x)_pid*, *pno*, *asame*, *new*, *newn*, *rstat*, *sex*, *birthm*, *birthy*, *age*, *agecat*, *primary*, *hhmem(x)*, *capi(x)*, *hssize*, *leavers*, *leftre*, *lefttm*, *lefty*, *lcounty*, *lcont*, *lmstat*, *lwith*, *leduc* a *hchange*. Ostatní proměnné z tohoto dotazníku nabývají pro bývalé členy hodnoty filtračních hodnot (97 / 999 997). V proměnných z ostatních instrumentů mají bývalí členové prázdné hodnoty (system missing), jsou jim však přiděleny hodnoty proměnných vztahujících se k průběhu a podmínkám dotazování. V proměnných k indikaci vyplněných dotazníků (*hh_hh*, *hh_indi* atd.) mají bývalí členové hodnotu 0 a nejsou pro ně spočteny váhy.

Počty nových a bývalých členů v domácnostech z jednotlivých let výzkumu jsou zaznamenány v *tabulce 25*. Údaje ve sloupci „domácnosti“ značí, v kolika domácnostech se vyskytuje alespoň jeden člen příslušného typu. Údaje za bývalé členy jsou uvedeny jak včetně duplicitních osob (tedy členů rozdělených domácností, kteří se díky vyšetření více než jedné dceřiné domácnosti vyskytují v souboru dvakrát), tak bez nich.

Tabulka 25: počty nových a bývalých členů domácností

	W2		W3		W4	
	domácnosti	jednotlivci	domácnosti	jednotlivci	domácnosti	jednotlivci
noví členové	235	258	234	278	175	204
bývalí členové včetně duplicitní osob	224	282	200	240	160	206
bývalí členové bez duplicitních osob	192	231	178	205	144	166
duplicitní osoby	32	51	22	35	17	40

8. Proměnné v podmíněném dotazování

S cílem usnadnit respondentům odpovídání, zvýšit efektivitu dotazování a omezit chybovost odpovědí byly CAPI domácnostní a individuální dotazníky ve druhé až čtvrté vlně uzpůsobeny tak, aby braly v úvahu charakteristiky osob a domácností z minulých ročníků (tzv. podmíněné dotazování, „dependent interviewing“).

Protože CAPI individuálním dotazníkem procházel respondent s ohledem na to, zda tento typ instrumentu již někdy v minulosti vyplnil, byla připravena sada indikátorů k rozlišení původních a nových respondentů. Proměnná *capi(x)* určuje, zda osoba vyplnila CAPI individuální dotazník v ročníku, který odpovídá číselnému indexu proměnné. V datovém souboru z příslušného ročníku jsou obsaženy vždy jen indikace vztahující se k přechozím vlnám (v aktuální vlně značí respondenty CAPI individuálního dotazníku proměnná *indi*). Ve třetí a čtvrté vlně byla vytvořena proměnná *capitat*, jejíž hodnoty odpovídají různým kombinacím vyplnění CAPI individuálního dotazníku v předchozích vlnách. Pro jednoduchou formulaci filtračních podmínek byla proměnná *capitat* dichotomizována do proměnné *capi*, která určuje, zda respondent vyplnil v minulosti CAPI individuální dotazník alespoň jednou.

V CAPI domácnostním a individuálním dotazníku se využívaly vybrané proměnné z předchozích let následujícími způsoby (typy využití jednotlivých proměnných shrnuje *tabulka 26*):

- 1 předvyplnění odpovědi hodnotou z minulého dotazování. Nová odpověď má být vybrána, pokud došlo ke změně nebo hodnota byla při minulém dotazování chybně zaznamenána,
- 2 připomenutí odpovědi z předchozího dotazování v textu otázky. Respondent v otázce uvádí, zda došlo či nedošlo ke změně stavu,
- 3 připomenutí odpovědi z předchozího dotazování v textu otázky. Respondent na otázku odpovídá, jaká je situace nyní,
- 4 hodnota byla využita při filtrování, nikoli připomenuta respondentovi.

Pokud se současný stav zjišťoval pomocí připomenutí minulé odpovědi (typ 2 a 3 v přehledu výše), byly při čištění dat do cílových proměnných doplněny hodnoty dle následujících pravidel. Pokud respondent ve srovnávací otázce (např. *wrkpsch*) uvedl, že nedošlo ke změně stavu (typ 2), byla do cílové proměnné (např. *wrkps*) doplněna minulé hodnota. Například pokud respondent ve druhém ročníku uvedl v proměnné *wrkpsch*, že jeho zaměstnání je stále stejné jako v prvním roce, bylo do proměnné *wrkps* doplněno zaměstnání z prvního roku. U otázek, kde respondent po připomenutí odpovědi uváděl aktuální stav, nikoli srovnání s minulým stavem (typ 3), byl do cílové proměnné (např. *svise*) nahrán údaj o současném stavu ze srovnávací proměnné (např. *svisech*). Například pokud ve druhé vlně respondent odpovídal na srovnávací otázku *svisech*, byla tato odpověď převedena i do cílové proměnné *svise*. Konkrétní způsob doplnění hodnot je uveden v dotaznících u příslušných proměnných.

Pozor, u otázek, které vůbec nebyly pokládány původním respondentům CAPI individuálního dotazníku (např. *eduprim*, *edusec*), hodnoty dodatečně vkládány nebyly. Hodnoty všech těchto proměnných jsou pro uvedené respondenty dostupné jen v datovém souboru z ročníku, v němž respondent poprvé vyplnil CAPI individuální dotazník.

Zatímco při podmíněném dotazování ve druhém ročníku mohly být využity jednoduše hodnoty z první vlny, ve třetím a čtvrtém ročníku bylo třeba počítat s tím, že potřebné hodnoty pro danou domácnost či osobu nemusí být v bezprostředně předcházející vlně dostupné. Zejména v případě hodnot z CAPI individuálního dotazníku bylo třeba sáhnout nikoli k minulé vlně, ale k poslední vlně, v níž respondent tento instrument vyplnil. Ve třetím a čtvrtém ročníku byly proto pro účely podmíněného dotazování připraveny pomocné proměnné obsahující pro danou domácnost či osobu hodnotu z poslední relevantní vlny.

Do těchto pomocných proměnných byly nejčastěji vkládány hodnoty z minulé vlny (proměnné z CAPI dotazníku pro domácnosti) nebo z poslední vlny, kdy respondent vyplnil CAPI individuální dotazník (proměnné z CAPI individuálního dotazníku). V několika případech však byly do pomocné proměnné vloženy hodnoty z dřívějších vln. Do pomocných proměnných *o_mstat*, *o_educ* a *o_rel(x)* byly nahrány hodnoty z vlny, v níž daná osoba naposledy patřila mezi členy domácnosti (tedy neměla status odchozího člena *rstat* = 5). Pro většinu osob to znamená minulou vlnu, pro bývalé členy však ročník předtím, než z domácnosti odešli. U pomocných proměnných *o_dwelob*, *o_faliv*, *o_maliv* a *o_fowny* nebylo možné pracovat s hodnotou z minulého ročníku (domácnost) či posledního CAPI individuálního dotazníku, protože u nich neprobíhalo v jednotlivých letech průběžné doplňování hodnot (popsáno výše). Do pomocných proměnných proto byla nahrána odpověď z roku posledního položení dané otázky.

V rámci přípravy dat z předchozích vln pro využití při podmíněném dotazování byly začištěny textové proměnné *sname1* (standardizace různě uváděných názvů škol včetně doplnění ulice a města z proměnných *sstr1* a *smun1*), *sprog1* a *wrkps* (oprava pravopisných chyb a překlepů).

Tabulka 26: proměnné vztahující se k podmíněnému dotazování

(cílová) proměnná	typ využití	rovnávací proměnná	
capi(x)	-	-	indikace, zda člen domácnosti vyplnil CAPI INDI dotazník ve vlně (x)
capistat	-	-	ukazatel vyplnění CAPI INDI v předchozích letech
capi	-	-	ukazatel vyplnění CAPI INDI v alespoň jednom z předchozích ročníků
name	1	-	křestní jméno člena domácnosti
sex	1	-	pohlaví člena domácnosti
birthm	1	-	měsíc narození člena domácnosti
birthy	1	-	rok narození člena domácnosti
mstat	4	-	rodinný stav člena domácnosti
educ	1	-	nejvyšší dosažené vzdělání člena domácnosti
rel(x)	1	-	vztah člena domácnosti k osobám $pno = x$
htnr	2	htnrch	právní důvod užívání bydlení
dwelob	4	-	
sfield	4	-	
stype	4	-	
sname1 *	2	snamech1	název školy studované v současnosti
sprog1 *	2	sprog1ch	obor studovaný v současnosti
wrkps *	2	wrkpsch	pracovní činnost respondenta
isco08	2	wrkpsch	pracovní činnost respondenta zakódovaná dle klasifikace CZ-ISCO
svise	3	svisech	zda má respondent zodpovědnost za vedení jiných zaměstnanců
suborch	2	subor	počet podřízených respondenta
empl	3	emplch	zda má respondent zaměstnance
emplsch	2	emplsch	počet zaměstnanců respondenta
fsizech	2	fsizech	počet zaměstnanců ve firmě, kde respondent pracuje
contr	2	contrch	typ pracovní smlouvy
faliv	4	-	otec osoby je naživu
maliv	4	-	matka osoby je naživu
fowny	4	-	rok, kdy se osoba stala poprvé vlastníkem nemovitosti

*¹) Z důvodu anonymizace dat nezařazeno do datového souboru.

9. Generované proměnné

Datový soubor obsahuje několik proměnných, jejichž hodnoty jsou odvozeny z jiných proměnných. Některé z těchto proměnných byly generovány v průběhu dotazování a sloužily k filtrování v následných otázkách, ostatní byly zkonstruovány ex-post.

Generované proměnné na úrovni domácnosti mají shodnou hodnotu pro všechny členy v domácnosti. Individuální generované proměnné odvozené z proměnných v CAPI dotazníku pro domácnost mají hodnotu pro všechny členy domácnosti. Výjimkou jsou bývalí členové domácnosti, kterým je přidělena platná hodnota jen u vybraných proměnných obou těchto typů (podrobnosti v oddílu 7. *Proměnné popisující změnu složení domácnosti*).

Individuální generované proměnné odvozené z proměnných v CAPI individuálním nebo proxy dotazníku mají hodnoty pro respondenty, kteří vyplnili daný instrument v příslušném roce (výjimka u proměnné *sumchild* je popsána v *tabulce 29*).

Tabulka 27: generované proměnné na úrovni domácnosti (CAPI dotazník pro domácnost)

hhsiz	počet osob v domácnosti
nborn	počet osob v domácnosti ve věku do 1 roku ($age < 1$)
ch0to5	počet osob v domácnosti ve věku do 5 let včetně ($age \leq 5$)
ch6to11	počet osob v domácnosti ve věku od 6 do 11 let včetně ($6 \geq age \leq 11$)
ch10to17	počet osob v domácnosti ve věku od 10 do 17 let včetně ($10 \geq age \leq 17$)
ch0to17	počet osob v domácnosti ve věku do 17 let včetně ($age \leq 17$)
p0to26	počet osob v domácnosti ve věku do 26 let včetně ($age \leq 26$)
adults	počet dospělých osob v domácnosti ($age \geq 18$)
new	přítomnost nových členů v domácnosti
newn	počet nových členů v domácnosti ($hhmem1 = 2 / w1_pid = 97$)
leavers	počet bývalých členů domácnosti ($rstat = 5$)
hchange	indikace změny ve složení domácnosti (tj. přítomnost nových nebo bývalých členů domácnosti)
nown	zda se domácnost stala od minulého dotazování vlastníkem bydlení

CAPI dotazník pro domácnost obsahuje od druhé vlny proměnnou *#hhsiz*, která odpovídá počtu všech osob zahrnutých pod danou domácnost, tedy včetně bývalých členů (kromě zemřelých). Jedná se pouze o pomocnou proměnnou využitou při dotazování pro práci s bývalými členy domácnosti, a není proto zařazena do datového souboru.

Na úrovni domácnosti bylo zkonstruováno několik ukazatelů struktury domácnosti, a to proměnná *hhtype* a sada proměnných *hhtype2* až *hhtype6* (tabulky 28 a 29). Proměnné *hhtype2* až *hhtype6* značí přítomnost daných skupin osob v domácnosti bez ohledu na přítomnost jakýchkoli dalších osob, zatímco kategorie proměnné *hhtype* značí přítomnost právě jen popsaných skupin osob. Například manželský pár žijící s matkou manžela (tchyně manželky) bude mít v *hhtype* hodnotu 7 (ostatní) a hodnotu 1 (ano) v proměnných *hhtype2* a *hhtype3*.

Tabulka 28: ukazatele přítomnosti skupin osob v domácnosti

hhtype2	zda je v domácnosti přítomná (alespoň jedna) skupina osob splňující podmínky: dvě osoby v domácnosti jsou vzájemně manželé (<i>rel</i> = 1), stálí partneři (<i>rel</i> = 2) nebo registrovaní partneři (<i>rel</i> = 3)
hhtype3	zda je v domácnosti přítomná (alespoň jedna) skupina osob splňující podmínky: jedna osoba, která nemá v domácnosti manžela/manželku (<i>rel</i> ≠ 1), stálého partnera/stálou partnerku (<i>rel</i> ≠ 2), registrovaného partnera/registrovaného partnerku (<i>rel</i> ≠ 3), je biologický (<i>rel</i> = 9), nevlastní (<i>rel</i> = 10), adoptivní rodič (<i>rel</i> = 11) nebo pěstoun (<i>rel</i> = 12) alespoň jedné další osoby, z nichž žádná není mladší 18 let (<i>age</i> ≥ 18)
hhtype4	zda je v domácnosti přítomná (alespoň jedna) skupina osob splňující podmínky: jedna osoba, která nemá v domácnosti manžela/manželku (<i>rel</i> ≠ 1), stálého partnera/stálou partnerku (<i>rel</i> ≠ 2), registrovaného partnera/registrovaného partnerku (<i>rel</i> ≠ 3), je biologický (<i>rel</i> = 9), nevlastní (<i>rel</i> = 10), adoptivní rodič (<i>rel</i> = 11) nebo pěstoun (<i>rel</i> = 12) alespoň jedné další osoby, z nichž nejméně jedna je mladší 18 let (<i>age</i> <18)
hhtype5	zda je v domácnosti přítomná (alespoň jedna) skupina osob splňující podmínky: dvě osoby v domácnosti, které jsou vzájemně manželé (<i>rel</i> = 1), stálí partneři (<i>rel</i> = 2) nebo registrovaní partneři (<i>rel</i> = 3), jsou biologičtí (<i>rel</i> = 9), nevlastní (<i>rel</i> = 10), adoptivní rodiče (<i>rel</i> = 11) nebo pěstouni (<i>rel</i> = 12) alespoň jedné další osoby v domácnosti, z nichž žádná není mladší 18 let (<i>age</i> ≥ 18)
hhtype6	zda je v domácnosti přítomná (alespoň jedna) skupina osob splňující podmínky: dvě osoby v domácnosti, které jsou vzájemně manželé (<i>rel</i> = 1), stálí partneři (<i>rel</i> = 2) nebo registrovaní partneři (<i>rel</i> = 3), jsou biologičtí (<i>rel</i> = 9), nevlastní (<i>rel</i> = 10), adoptivní rodiče (<i>rel</i> = 11) nebo pěstouni (<i>rel</i> = 12) alespoň jedné další osoby v domácnosti, z nichž alespoň jedna je mladší 18 let (<i>age</i> <18)

Tabulka 29: typ domácnosti podle struktury

1	domácnost jednotlivce	<ul style="list-style-type: none"> domácnost má jednoho člena (<i>hhs</i> = 1)
2	partnerská/manželská domácnost	<ul style="list-style-type: none"> domácnost má dva členy (<i>hhs</i> = 2), členové domácnosti jsou vzájemně manželé (<i>rel</i> = 1), stálí partneři (<i>rel</i> = 2) nebo registrovaní partneři (<i>rel</i> = 3)
3	rodič s dospělým dítětem/dospělými dětmi	<ul style="list-style-type: none"> domácnost má alespoň dva členy (<i>hhs</i> ≥ 2), jeden člen domácnosti je biologický (<i>rel</i> = 9), nevlastní (<i>rel</i> = 10), adoptivní (<i>rel</i> = 11) rodič nebo pěstoun (<i>rel</i> = 12) všech ostatních, ostatní členové jsou vzájemně biologičtí (<i>rel</i> = 13), nevlastní (<i>rel</i> = 14, <i>rel</i> = 15), adoptivní (<i>rel</i> = 16) sourozenci nebo sourozenci v pěstounské péči (<i>rel</i> = 17), žádný z ostatních členů (dětí) není mladší 18 let (<i>age</i> ≥ 18)
4	rodič s alespoň jedním dítětem mladším 18 let	<ul style="list-style-type: none"> domácnost má alespoň dva členy (<i>hhs</i> ≥ 2), jeden člen domácnosti je biologický (<i>rel</i> = 9), nevlastní (<i>rel</i> = 10), adoptivní (<i>rel</i> = 11) rodič nebo pěstoun (<i>rel</i> = 12) všech ostatních, ostatní členové jsou vzájemně biologičtí (<i>rel</i> = 13), nevlastní (<i>rel</i> = 14, <i>rel</i> = 15), adoptivní (<i>rel</i> = 16) sourozenci nebo sourozenci v pěstounské péči (<i>rel</i> = 17), alespoň jeden z ostatních členů (dětí) je mladší 18 let (<i>age</i> < 18)
5	rodiče s dospělým dítětem/dospělými dětmi	<ul style="list-style-type: none"> domácnost má alespoň tři členy (<i>hhs</i> ≥ 3), dva členové jsou vzájemně manželé (<i>rel</i> = 1), stálí partneři (<i>rel</i> = 2) nebo registrovaní partneři (<i>rel</i> = 3), manželé/partneři jsou biologičtí (<i>rel</i> = 9), nevlastní (<i>rel</i> = 10), adoptivní (<i>rel</i> = 11) rodiče nebo pěstouni (<i>rel</i> = 12) všech ostatních, ostatní členové jsou vzájemně biologičtí (<i>rel</i> = 13), nevlastní (<i>rel</i> = 14, <i>rel</i> = 15), adoptivní (<i>rel</i> = 16) sourozenci nebo sourozenci v pěstounské péči (<i>rel</i> = 17), žádný z ostatních členů (dětí) není mladší 18 let (<i>age</i> ≥ 18)
6	rodiče s alespoň jedním dítětem mladším 18 let	<ul style="list-style-type: none"> domácnost má alespoň tři členy (<i>hhs</i> ≥ 3), dva členové jsou vzájemně manželé (<i>rel</i> = 1), stálí partneři (<i>rel</i> = 2) nebo registrovaní partneři (<i>rel</i> = 3), manželé/partneři jsou zároveň biologičtí (<i>rel</i> = 9), nevlastní (<i>rel</i> = 10), adoptivní (<i>rel</i> = 11) rodiče nebo pěstouni (<i>rel</i> = 12) všech ostatních, ostatní členové jsou vzájemně biologičtí (<i>rel</i> = 13), nevlastní (<i>rel</i> = 14, <i>rel</i> = 15), adoptivní (<i>rel</i> = 16) sourozenci nebo sourozenci v pěstounské péči (<i>rel</i> = 17), alespoň jeden z ostatních členů (dětí) je mladší 18 let (<i>age</i> < 18)
7	ostatní	<ul style="list-style-type: none"> ostatní domácnosti
8	nekonzistenční vztahy v domácnosti	<ul style="list-style-type: none"> vztahy v domácnosti jsou vzájemně neslučitelné

Tabulka 30: generované proměnné na úrovni jednotlivce

age	věk v letech odvozený od měsíce a roku narození (<i>birthm, birthy</i>) a měsíce a roku vyplnění CAPI dotazníku pro domácnost (<i>hstartm, hstarty</i>). Pokud se věk uvedený v PAPI dotaznících a časových snímcích (<i>papiage, diaryage</i>) lišil o jeden rok a odpovídal věkové populaci příslušného instrumentu, byla do proměnné <i>age</i> nahrána hodnota <i>papiage/diaryage</i> .
agecat	věk kategorizovaný po 10 letech
primary	výskyt osoby ve vzorku: primární (člen nerozdělených domácností, aktuální člen rozdělené domácnosti) nebo sekundární (bývalý člen rozdělené domácnosti)
hhmem(x)	ukazatel členství osoby v domácnosti ve vlně (x)
capi(x)	indikace, zda člen domácnosti vyplnil CAPI INDI dotazník ve vlně (x)
capistat	ukazatel vyplnění CAPI INDI v předchozích letech
capi	ukazatel vyplnění CAPI INDI v alespoň jednom z předchozích ročníků
livesp^{*)}	zda v domácnosti žije manžel/manželka (<i>rel = 1</i>) nebo registrovaný partner/registrovaná partnerka (<i>rel = 3</i>) osoby
livewith^{*)}	zda v domácnosti žije stálý partner/stálá partnerka (<i>rel = 2</i>) osoby
livefa^{*) **)}	zda v domácnosti žije muž (<i>sex = 1</i>), který je biologický (<i>rel = 9</i>), nevlastní (<i>rel = 10</i>), adoptivní otec (<i>rel = 11</i>) nebo pěstoun (<i>rel = 12</i>) osoby
livo^{*) **)}	zda v domácnosti žije žena (<i>sex = 2</i>), která je biologická (<i>rel = 9</i>), nevlastní (<i>rel = 10</i>), adoptivní matka (<i>rel = 11</i>) nebo pěstounka (<i>rel = 12</i>) osoby
livebifa^{*)}	zda v domácnosti žije muž (<i>sex = 1</i>), který je biologický otec (<i>rel = 9</i>) osoby
livebimo^{*)}	zda v domácnosti žije žena (<i>sex = 2</i>), která je biologická matka (<i>rel = 9</i>) osoby
livegfa^{*)}	zda v domácnosti žije muž (<i>sex = 1</i>), který je dědeček osoby (<i>rel = 21</i>)
livegmo^{*)}	zda v domácnosti žije žena (<i>sex = 2</i>), která je babička osoby (<i>rel = 21</i>)
bipar^{*)}	zda alespoň jeden člen domácnosti je biologické dítě (<i>rel = 4</i>) osoby
child	počet členů domácnosti, kteří jsou biologické (<i>rel = 4</i>), nevlastní (<i>rel = 5</i>) nebo adoptivní (<i>rel = 6</i>) děti osoby nebo jeho/její děti v pěstounské péči (<i>rel = 7</i>)
childy	počet členů domácnosti mladších 18 let (<i>age < 18</i>), kteří jsou biologické (<i>rel = 4</i>), nevlastní (<i>rel = 5</i>) nebo adoptivní (<i>rel = 6</i>) děti osoby nebo jeho/její děti v pěstounské péči (<i>rel = 7</i>)
childa	počet dospělých (<i>age ≥ 18</i>) členů domácnosti, kteří jsou biologické (<i>rel = 4</i>), nevlastní (<i>rel = 5</i>) nebo adoptivní (<i>rel = 6</i>) děti osoby nebo jeho/její děti v pěstounské péči (<i>rel = 7</i>)
etran	změna ekonomického postavení mezi vlnami
relup	změna rezidenčního partnera mezi vlnami
own	zda člen domácnosti patří mezi vlastníky bytu/domu, kde domácnost bydlí
coo	zda člen domácnosti patří mezi členy družstva vlastního byt, kde domácnost bydlí
part	zda má respondent CAPI individuálního dotazníku rezidenčního (<i>livesp = 1</i> or <i>livewith = 1</i>) nebo nerezidenčního partnera (<i>date = 1</i>) nebo je ženatý/vdaná či v registrovaném partnerství (<i>mstat = 1</i> or <i>mstat = 2</i>). Pokud nebylo možné hodnotu stanovit z důvodu neplatných odpovědí na zdrojových proměnných, byla do proměnné <i>part</i> nahrána platná odpověď z otázky <i>papipart</i> v PAPI dotazníku pro dospělé.
sumchild	celkový počet biologických (<i>rel = 4</i>) a adoptivních (<i>rel = 6</i>) rezidenčních a nerezidenčních (<i>nrechn</i>) dětí respondentů CAPI individuálního a proxy dotazníku. Ve třetím ročníku byla

proměnná *nrechn* zařazena do CAPI individuálního dotazníku pouze těm respondentům, kteří na tento instrument dosud nikdy neodpovídali (*capi1* = 2 and *capi2* = 2). Proměnná *sumchild* je ve třetím ročníku proto vypočítána z hodnoty *nrechn* získané v roce, v němž respondent naposledy vyplnil CAPI individuální dotazník.

nage(x) věk nerezidenčních dětí odvozený z proměnných *nbirthm(x)* a *nbirthy(x)*

mrenum počet zvolených důvodů k přestěhování (*mwork*, *mfam*, *mhous*, *melse*) respondentů CAPI individuálního dotazníku

*¹ Proměnné *live-* a *bipar* jsou odvozeny ze vztahů mezi členy domácnosti (proměnné *rel-*) a nabývají pouze hodnot 1 (ano) a 2 (ne). Tyto proměnné tedy nevypovídají o tom, zda mimo domácnost žije někdo, kdo má k osobě daný vztah (např. hodnota „ne“ v proměnné *livesp* znamená, že nikdo v domácnosti není manžel/manželka osoby, nikoli že osoba má manžel/manželku, který/která žije mimo domácnost).

**¹ U osob, které mají v domácnosti nevlastního rodiče stejného pohlaví jako pohlaví biologického rodiče, nabývá hodnoty „ano“ také proměnná vážící se k druhému pohlaví, než je pohlaví biologického rodiče.

10. Stratifikační postavení

Hodnotám proměnných *isco08*, *fisco08*, *misco08*, *aisco08* a *eisco08* byly přiřazeny hodnoty ukazatele ISEI (International Socio-Economic Index)⁷.

Proměnné *estat* (*festat*, *mestat*), *jobtp*, *svise* (*fsvise*, *msvise*), *empl* (*fempl*, *mempl*), *empls* (*fempls*, *mempls*) a *isco08* (*fisco08*, *misco08*) byly použity ke konstrukci ukazatele ESeC (European Socio-economic Classification)⁸.

Ukazatele ISEI jsou dostupné pro respondenty CAPI individuálního dotazníku, osoby, za něž byl vyplněn CAPI proxy dotazník, a respondenty PAPI dotazníků pro děti. Ukazatele ESeC jsou dostupné pro respondenty CAPI individuálního dotazníku a osoby, za něž byl vyplněn CAPI proxy dotazník.

SPSS syntax použitá ke konstrukci ukazatelů ISEI a ESeC je dostupná jako samostatný soubor.

Tabulka 31: ukazatele stratifikačního postavení

ISEIisco08	socioekonomický index respondenta
ISEIfisco08	socioekonomický index otce respondenta (v době, kdy respondentovi bylo 16 let)
ISEImisco08	socioekonomický index matky respondenta (v době, kdy respondentovi bylo 16 let)
ISEIaisco08	socioekonomický index práce, kterou chce dítě v budoucnu vykonávat
ISEIeisco08	socioekonomický index práce, kterou si dítě myslí, že bude v budoucnu vykonávat
ESeC08	ESeC 9 tříd respondent
ESeC08f	ESeC 9 tříd otec respondenta (v době, kdy respondentovi bylo 16 let)
ESeC08m	ESeC 9 tříd matka respondenta (v době, kdy respondentovi bylo 16 let)
ESeC08_6	ESeC 6 tříd respondent
ESeC08f_6	ESeC 6 tříd otec respondenta (v době, kdy respondentovi bylo 16 let)
ESeC08m_6	ESeC 6 tříd matka respondenta (v době, kdy respondentovi bylo 16 let)

⁷ 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>.

⁸ Harrison, Eric. „European Socio-economic Classification.“ <http://ekharrison.weebly.com/european-socio-economic-classification-esec.html> <February 5 2015>.

11. Proměnné individuálního příjmu

Z otázek vztahujících se k příjmu jednotlivce z různých zdrojů (CAPI individuální a proxy dotazník) bylo zkonstruováno několik shrnujících proměnných, a to příjem ze závislé činnosti, příjem z podnikání, příjem ze sociálních benefitů, příjem z důchodů a příjem z kapitálového majetku včetně dalších příjmů.

Všechny uvedené kategorie příjmu jsou vyjádřeny jako měsíční částky. Zatímco u příjmů ze závislé činnosti, důchodů a sociálních benefitů odpovídá hodnota skutečné měsíční částce (měsíční částka uvedená respondentem nebo přepočet roční částky na počet měsíců pobírání příjmu), u příjmů z podnikání a kapitálových příjmů se jedná o dvanáctinu roční částky uvedené respondentem.

Pokud respondent daný příjem nepobíral, byla mu v odpovídající generované proměnné přiřazena hodnota 0. V případech, kdy nebylo možné proměnnou zkonstruovat z důvodu neplatné odpovědi na alespoň jedné zdrojové proměnné, obsahují data hodnotu 98 / 999 998.

Příjmy ze závislé činnosti jsou vyjádřeny pro respondenty CAPI individuálního dotazníku i členy domácnosti, za něž byl proveden proxy rozhovor. Všechny ostatní příjmové proměnné mohou být vypočteny jen pro respondenty CAPI individuálního dotazníku.

Hodnoty **příjmů ze závislé činnosti** jsou dostupné jak ve formě částek (proměnné *dignum* a *dinum*), tak ve formě kategorií pokrývajících příjmové intervaly (proměnné *digcam* a *dincam*). Respondenti při dotazování však uváděli jen jeden typ údajů, druhý byl odvozen při konstrukci generovaných proměnných. Respondenti CAPI individuálního dotazníku byli přednostně dotazováni na částky a na otázku po příjmové kategorii odpovídali jen v případě, že částku odmítli uvést nebo tuto informaci neznali. V proxy rozhovorech byly dotazovány jen příjmové kategorie. Částky jsou z příjmových kategorií vyjádřeny jako střední hodnoty intervalu v dané příjmové kategorii (např. u kategorie 1 až 4 999 Kč je střední hodnotou intervalu 2 500 Kč) s výjimkou vrchních kategorií, kde přiřazená částka odpovídá mediánu číselných hodnot uvedených respondenty spadajících do dané kategorie.

Proměnné *c_dignum* a *c_dinum* indikují zdrojový údaj pro výpočet proměnných *dignum* a *dinum*.

Tabulka 32: způsob konstrukce proměnných *dignum* a *dinum*

hodnota	dotazník	popis konstrukce
0	CAPI individuální	v proměnné <i>payg/payn</i> uvedena hodnota 0
1	CAPI individuální	v proměnné <i>payg/payn</i> uvedena měsíční částka (<i>paygam</i> = 2)
2	CAPI individuální	v proměnné <i>payg/payn</i> uvedena roční částka (<i>paygam</i> = 1), měsíční částka vypočtena jako podíl ročního příjmu a počtu měsíců pobírání v proměnné <i>paymon</i>
3	CAPI individuální	neznalost odpovědi či odmítnutí v otázce <i>payg/payn</i> , avšak v otázce <i>paygcat/payncat</i> uvedena platná odpověď. Dosazena střední hodnota příjmového pásma ve vybrané kategorii, u vrchního intervalu dosazen median částek příjmů spadajících do daného intervalu.
4	CAPI proxy	dotázána pouze otázka na příjmové kategorie <i>paygcat/payncat</i> . Dosazena střední hodnota příjmového pásma ve vybrané kategorii, u vrchního intervalu dosazen median částek příjmů spadajících do daného intervalu.

Proměnné *signum* a *sinum* vyjadřují měsíční hrubé a čisté **příjmy z podnikání**. Protože k příjmům z podnikání není dotazován počet měsíců pobírání příjmu, je měsíční částka vypočtena jako jedna dvanáctina ročního příjmu. Kladné částky představují zisk z podnikání, záporné částky ztrátu.

Příjmy z důchodů obsahuje proměnná *penum*. Zahrnuty jsou důchody starobní, vdovské/vdovecké/sirotčí a invalidní. Hodnota odpovídá součtu měsíčních částek ve zdrojových proměnných.

Součet měsíčních částek jednotlivých dotazovaných **sociálních benefitů** (podpora v nezaměstnanosti, dávky v hmotné nouzi, nemocenská, peněžitá pomoc v mateřství, rodičovský příspěvek a stipendia) představuje proměnná *benum*.

Proměnná *cognum* je součtem **příjmů z kapitálových majetků a jakýchkoli dalších příjmů** nezahrnutých do předchozích kategorií. Měsíční částka je vypočtena jako jedna dvanáctina ročního příjmu.

Tabulka 33: proměnné individuálního příjmu

		dotazník	zdrojové proměnné
dignum	částka hrubého měsíčního příjmu ze závislé činnosti v Kč	CAPI individuální a proxy	<i>payg, paygam, paymon, paygcat</i>
c_dignum	výchozí údaj pro konstrukci proměnné dignum	CAPI individuální a proxy	<i>payg, paygam, paymon, paygcat</i>
dinum	částka čistého měsíčního příjmu ze závislé činnosti v Kč	CAPI individuální a proxy	<i>payn, paynam, paymon, payncat</i>
c_dinum	výchozí údaj pro konstrukci proměnné dinum	CAPI individuální a proxy	<i>payn, paynam, paymon, payncat</i>
digcam	kategorie hrubého měsíčního příjmu ze závislé činnosti v Kč	CAPI individuální a proxy	<i>payg, paygam, paymon, paygcat</i>
dincam	kategorie čistého měsíčního příjmu ze závislé činnosti v Kč	CAPI individuální a proxy	<i>payn, paynam, paymon, payncat</i>
signum	částka hrubého měsíčního příjmu z podnikání v Kč (1/12 roční částky)	CAPI individuální	<i>sfpayg, gprof</i>
sinum	částka čistého měsíčního příjmu z podnikání v Kč (1/12 roční částky)	CAPI individuální	<i>sfpayn, nprof</i>
penum	částka měsíčního příjmu z důchodů (starobní, vdovský/vdovecký/sirotčí, invalidní) v Kč	CAPI individuální	<i>opena, wpena, dpena</i>
benum	částka měsíčního příjmu ze sociálních benefitů (podpora v nezaměstnanosti, dávky v hmotné nouzi, nemocenská, peněžité pomoci v mateřství, rodičovský příspěvek, stipendia) v Kč	CAPI individuální	<i>ucoma, nebena, sibena, mabena, pallowa, schola</i>
cognum	částka hrubého měsíčního příjmu z kapitálových majetků a dalších příjmů v Kč (1/12 roční částky)	CAPI individuální	<i>capina, othina</i>

12. Proměnné vztahující se k okolnostem dotazování

Datové soubory obsahují sadu údajů vypovídajících o kontextu a průběhu dotazování (tzv. paradata), které lze rozčlenit do čtyř skupin:

- místo bydliště domácnosti,
- charakteristiky dotazování,
- datum dotazování,
- vyplnění dotazníků v rámci domácnosti.

Místo bydliště domácnosti je v datech specifikováno na úrovni krajů (NUTS 3) a okresů (LAU 1). Zahrnuty jsou rovněž proměnné *strata* a *psu* značící jednotky využívané při výběru vzorku. *Strata* (1 až 58) představuje kombinaci velikosti místa bydliště (*msize*) a kraje (*region*). *Psu* odpovídá skupinám adres, které tvoří bytové jednotky z prvního kroku výběru a jim geograficky blízké adresy z druhého kroku výběru (postup výběru popsán v oddíle 2.3. *Výběr vzorku domácností*).

Velikost místa bydliště (*msize*), kraj (*region*) a okres (*county*) odpovídají adrese v době dotazování, zatímco *strata* a výběrový bod vycházejí z adresy v době provedení výběru (červen 2015). V prvním ročníku se však několik oslovených domácností před provedením rozhovorů odstěhovalo a byly dotázány ve svém novém bydlišti. Několik domácností má proto jinou lokalitu bydliště než ostatní domácnosti v daném výběrovém bodě či stratu.

Pokud se domácnost nepřestěhovala, je u ní ve všech vlnách zachována stejná velikost místa bydliště odpovídající počtům obyvatel v roce 2015, ačkoli podle oficiálních statistik se kategorie velikosti obce změnila.

Tabulka 34: místo bydliště domácnosti

msize	velikost obce bydliště
region	kraj (NUTS 3)
county	okres (LAU 1)
strata	strata (1 až 58), pouze vlna 1
psu	výběrový bod (1 až 1275), pouze vlna 1

Data obsahují indikátory umožňující rozlišit mód dotazování (CAPI = osobní rozhovor s tazatelem a vyplnění papírových dotazníků, CAWI = vyplnění všech instrumentů přes internet), agenturu, která dotázala danou domácnost, včetně tazatele. Číslo tazatele je uvedeno pouze v případě, že daná domácnost byla vyšetřena osobním rozhovorem, nikoli prostřednictvím CAWI. Ukazatel *intid* odpovídá tazatelským číslům obsaženým v datech z kontaktních formulářů a souborech s údaji o tazatelích.

Doplněno je rovněž několik indikací typu vyplněných instrumentů, tazatelem zaznamenaný důvod provedení proxy rozhovoru a odvozený vztah mezi proxy respondentem a osobou, za niž vypovídá.

Tabulka 35: charakteristiky dotazování

qmode	zda se domácnost zúčastnila osobního rozhovoru nebo prostřednictvím CAWI
agen	agentura, která provedla dotazování domácnosti
intid	číslo tazatele (tentýž tazatel má stejné číslo v datech ze všech let výzkumu)
iorder	pořadí vyplnění CAPI individuální dotazníku v domácnosti (pouze pro respondenty, s nimiž byl přímo vyplněn CAPI individuální dotazník, nikoli formou proxy)
proxy	zda bylo CAPI individuální dotazování provedeno jen formou zkráceného proxy rozhovoru
relpro	vztah proxy respondenta a osoby, za niž dotazník vyplňuje
reaspro	důvod provedení proxy rozhovoru
papi	druh vyplněného PAPI dotazníku (pouze pro respondenty, kteří vyplnili PAPI dotazník)

Dny a měsíce začátků a ukončení CAPI dotazování automaticky zaznamenával CAPI dotazový software. V PAPI dotaznících vyplňovali údaje sami respondenti.

Tabulka 36: datum dotazování

hstartm	měsíc začátku CAPI dotazníku pro domácnost
hstartd	den začátku CAPI dotazníku pro domácnost
hendm	měsíc ukončení CAPI dotazníku pro domácnost
hendd	den ukončení CAPI dotazníku pro domácnost
istartm	měsíc začátku CAPI individuálního dotazníku
istartd	den začátku CAPI individuálního dotazníku
iendm	měsíc ukončení CAPI individuálního dotazníku
iendd	den ukončení CAPI individuálního dotazníku
pstartmo	měsíc začátku CAPI proxy dotazníku
pstartd	den začátku CAPI proxy dotazníku
pendm	měsíc ukončení CAPI proxy dotazníku
pendd	den ukončení CAPI proxy dotazníku
papimon	měsíc vyplňování PAPI dotazníku
papiday	den vyplňování PAPI dotazníku

V datovém souboru jsou za jednotlivé domácnosti spočteny údaje o množství vyplněných dotazníků. Podíly vyplněných instrumentů jsou vyjádřeny jako počet vyplněných instrumentů daného typu děleno počtem osob v daném věku v domácnosti. Pokud v domácnosti nebyly žádné děti ve věku 10 až 17 let, hodnoty proměnných vztahujících se k instrumentům pro děti jsou prázdné (system missing).

Tabulka 37: vyplnění dotazníků v domácnosti

h_indi	počet vyplněných CAPI individuálních dotazníků v domácnosti
hp_indi	podíl vyplněných CAPI individuálních dotazníků v domácnosti
h_indi_proxy	počet vyplněných CAPI individuálních a proxy dotazníků v domácnosti
hp_indi_proxy	podíl vyplněných CAPI individuálních a proxy dotazníků v domácnosti
h_papi_adult	počet vyplněných PAPI dotazníků pro dospělé v domácnosti
hp_papi_adult	podíl vyplněných PAPI dotazníků pro dospělé v domácnosti
h_papi_child	počet vyplněných PAPI dotazníků pro děti v domácnosti
hp_papi_child	podíl vyplněných PAPI dotazníků pro děti v domácnosti
h_diary_adult	počet vyplněných časových snímků pro dospělé v domácnosti
hp_diary_adult	podíl vyplněných časových snímků pro dospělé v domácnosti
h_diary_child	počet vyplněných časových snímků pro děti v domácnosti
hp_diary_child	podíl vyplněných časových snímků pro děti v domácnosti

13. Proměnné filtrované na základě hodnot jiné osoby

V CAPI individuálním dotazníku je zahrnuto několik otázek, které jsou filtrovány na základě vlastností jiného člena domácnosti, tj. filtruje se na základě hodnoty proměnné příslušející jinému členu domácnosti. Tyto otázky jsou pokládány rodičům, pokud jejich dítě v domácnosti splňuje dané podmínky, a odpověď na tyto otázky se váže k příslušnému dítěti.

Pokud filtr odkazuje k hodnotě, kterou proměnná nabývá pro člena domácnosti s *pno* x, je v kódu této proměnné využita předpona /x/. Uvedené značení je třeba odlišovat od přípony (x), která určuje konkrétní proměnnou z cyklu.

Například filtr pro proměnnou *imtalk2* „*rel2 = 4 | 5 | 6 | 7 AND 2age >= 10 AND 2age <= 17*“ znamená, že otázka bude položena, pokud je respondent rodičem 2. osoby v domácnosti a této osobě je mezi 10 a 17 lety.

Tabulka 38: proměnné filtrované na základě hodnot jiné osoby

imtalk(x), quar(x), hapschw(x)- hapliv(x)	pokládáno biologickým (<i>rel</i> = 4), nevlastním (<i>rel</i> = 5), adoptivním (<i>rel</i> = 6) rodičům nebo pěstounům (<i>rel</i> = 7), jejichž dítě je ve věku 10 až 17 let (<i>age</i> dítěte ≥ 10 a ≤ 17). Číslo v kódu proměnné označuje pořadí dítěte v domácnosti (<i>pno</i>).
chtime(x)	pokládáno biologickým (<i>rel</i> = 4), nevlastním (<i>rel</i> = 5), adoptivním (<i>rel</i> = 6) rodičům nebo pěstounům (<i>rel</i> = 7) dítěte. Číslo v kódu proměnné označuje pořadí dítěte v domácnosti (<i>pno</i>).
pdisch(x)	pokládáno biologickým (<i>rel</i> = 4), nevlastním (<i>rel</i> = 5), adoptivním (<i>rel</i> = 6) rodičům nebo pěstounům (<i>rel</i> = 7), jejichž dítě je starší 8 let (<i>age</i> dítěte ≥ 8). Číslo v kódu proměnné označuje pořadí dítěte v domácnosti (<i>pno</i>).
eduas(x)	pokládáno biologickým (<i>rel</i> = 4), nevlastním (<i>rel</i> = 5), adoptivním (<i>rel</i> = 6) rodičům nebo pěstounům (<i>rel</i> = 7), jejichž dítě je mladší 25 let (<i>age</i> dítěte < 25) a studuje (<i>std</i> dítěte = 1). Číslo v kódu proměnné označuje pořadí dítěte v domácnosti (<i>pno</i>).
eduex(x)	pokládáno biologickým (<i>rel</i> = 4), nevlastním (<i>rel</i> = 5), adoptivním (<i>rel</i> = 6) rodičům nebo pěstounům (<i>rel</i> = 7), jejichž dítě je mladší 25 let (<i>age</i> dítěte < 25) a studuje (<i>std</i> dítěte = 1). Číslo v kódu proměnné označuje pořadí dítěte v domácnosti (<i>pno</i>).
raidh(x)	pokládáno biologickým (<i>rel</i> = 4), nevlastním (<i>rel</i> = 5), adoptivním (<i>rel</i> = 6) rodičům nebo pěstounům (<i>rel</i> = 7), jejichž dítě je starší 15 let (<i>age</i> dítěte ≥ 15) a je vlastníkem bydlení domácnosti (<i>own</i> = 1) nebo členem družstva (<i>coo</i> = 1). Číslo v kódu proměnné označuje pořadí dítěte v domácnosti (<i>pno</i>).
rplaidh(x)	pokládáno biologickým (<i>rel</i> = 4), nevlastním (<i>rel</i> = 5), adoptivním (<i>rel</i> = 6) rodičům nebo pěstounům (<i>rel</i> = 7), jejichž dítě je starší 15 let (<i>age</i> dítěte ≥ 15), a pokud je vlastníkem současného bydlení (<i>own</i> = 1) nebo členem družstva (<i>coo</i> = 1), rodič mu dosud neposkytl finanční pomoc na pořízení bydlení (<i>raidh</i> = 4 88 99), nebo vlastníkem současného bydlení nebo členem družstva není (<i>raidh</i> = 97). Tato specifikace filtru platila pro 2. ročník, ve 4. byly podmínky upraveny. Číslo v kódu proměnné označuje pořadí dítěte v domácnosti (<i>pno</i>).

maben	pokládáno biologickým (<i>rel</i> = 4), nevlastním (<i>rel</i> = 5), adoptivním (<i>rel</i> = 6) rodičům nebo pěstounům (<i>rel</i> = 7), kteří mají v domácnosti alespoň jedno dítě do 2 let věku včetně (<i>age</i> dítěte <= 2)
pallow	pokládáno biologickým (<i>rel</i> = 4), nevlastním (<i>rel</i> = 5), adoptivním (<i>rel</i> = 6) rodičům nebo pěstounům (<i>rel</i> = 7), kteří mají v domácnosti alespoň jedno dítě do 4 let věku včetně (<i>age</i> dítěte <= 4)

14. Anonymizace dat

Instrumenty neobsahovaly otázky na osobní údaje respondentů. Osobní údaje (jméno včetně příjmení, kontaktní telefon a e-mail) byly získávány pouze pro účely opětovného kontaktu a zařazení domácností do slosování o výhry. Členové domácnosti poskytovali tyto údaje prostřednictvím odděleného formuláře, kde písemně souhlasili s využitím údajů jen pro uvedené účely. Kvůli propojitelnosti osobních údajů s odpověďmi z dotazníků byl ve čtvrté vlně (2018) zajišťován také písemný souhlas domácnosti s dotazováním a písemný souhlas rodičů s dotazováním dětí.

Do dat nebyly zařazeny vybrané proměnné, které poskytují detailnější informace o daném respondentovi (jejich přehled v oddílu 5.3 *Proměnné nezařazené do datového souboru*).

Číselné identifikátory domácností a jednotlivců (*hid*, *pid*) z primárních dat byly nahrazeny novými náhodnými identifikačními čísly, která se řídí pravidly konstrukce identifikátorů popsány v oddílu 6 *Identifikační proměnné*.

V hodnotách proměnných *nname1* až *nname8* (křestní jména nerezidenčních dětí) byla jména nahrazena číselnými kódy korespondujícími s různými křestními jmény.

15. Kontrola, zpracování a čištění dat

15.1. Kontroly dat

Do CAPI instrumentů bylo zakomponováno několik mechanismů na podporu validity a konzistence získávaných údajů. V CAPI dotazníku pro domácnost a individuálním dotazníku byly u vybraných otázek s číselnou odpovědí implementovány kontroly, které vyzvaly tazatele k ověření zadané číselné hodnoty, pokud se nacházela mimo určené rozmezí (upřesněno v dotaznících). V proměnných popisujících vztahy mezi členy domácnosti (*rel-*) byl po zadání vztahu první osoby ke druhé automaticky doplněn vztah druhé osoby k první. V úvodu CAPI individuálního dotazníku bylo ověřováno 8 základních sociodemografických údajů o respondentovi s možností opravy. Cílem bylo ověřit, zda je CAPI individuální dotazník vyplňován se správným respondentem, a zajistit správnost proměnných klíčových pro filtrování v dotazníku. Jak v CAPI dotaznících, tak v CAWI mutacích všech dotazníků nebylo možné nechat otázku bez odpovědi.

Tazatelé kontrolovali v domácnostech formální správnost vyplnění PAPI dotazníků a časových snímků a případně vyzvali respondenty k opravě. Dohlíželi také, aby v PAPI dotaznících a časových snímcích nezůstal nevyplněn velký počet otázek (resp. časových bloků).

CAPI data byla s týdenní (první vlna) či dvoutýdenní (druhá vlna) periodicitou exportována z dotazového software do formátu SPSS a realizátoři sběru dat i výzkumný tým prováděli průběžně kontroly konzistence mezi daty a dotazovými instrumenty, tedy zejména definic kódů jednotlivých odpovědí, dodržení povoleného rozsahu hodnot a nastavení filtrů. U numerických proměnných bez specifikovaného povoleného rozsahu hodnot byla provedena kontrola extrémních hodnot. Pokud se vyskytly, bylo zjišťováno, zda byly takto zadány při dotazování a nejsou důsledkem chybné manipulace s daty. Proměnné s odpovědí ve formě roku určité události (např. *madres*) prošly kontrolou konzistence výpovědi. Ačkoli byly během procesu kontroly identifikovány vzájemně nekonzistentní letopočty nebo nepravděpodobné numerické hodnoty, nedocházelo k jejich kumulaci u téhož respondenta do té míry, že by celek jeho odpovědí byl systematicky nelogický nebo svědčící o náhodném odpovídání. Hodnoty byly v těchto případech ponechány v nezměněné podobě.

Od druhého ročníku se kontroly zaměřovaly také na konzistenci základních sociodemografických údajů o členech šetřených domácností mezi jednotlivými vlnami. V případě značných nesrovnalostí (obvykle změna více než dvou charakteristik stabilního charakteru) byla domácnost zkontrolována s příslušným tazatelem (to bylo však možné jen u rozhovorů realizovaných formou CAPI) a vyloučena z dalšího zpracování, pokud změny nebyly uspokojivě vysvětleny.

Během monitorování terénních prací byla kontrolována shoda distribucí mezi agenturami zajišťujícími sběr dat. V případě rozdílů byly zjišťovány možné příčiny a zabezpečována harmonizace v postupu tazatelů. Kontrolou konzistence s dotazovými instrumenty, extrémních hodnot a shody distribucí prošla jak průběžná data, tak finální datový soubor.

Při pořizování dat z PAPI dotazníků (ruční pořizování v agentuře STEM/MARK, scanování v agentuře MEDIAN) a časových diářů (ruční pořizování v obou agenturách) byla kontrolována úplnost vyplnění a zároveň propojitelnost s CAPI instrumenty pomocí identifikačních čísel respondentů. V agentuře

STEM/MARK bylo ve všech vlnách pomocí dvojího opakovaného vstupu dat verifikováno 20 % datových souborů z PAPI dotazníků a časových diářů. Agentura Median prováděla u PAPI dotazníků kontroly nastavení pořizovací masky a manuální kontrolu dat pořízených z několika desítek náhodně vybraných dotazníků. Dále manuálně prověřila všechny záznamy vyhodnocené při scanování jako bez odpovědi a případné platné hodnoty doplnila ručně. V agentuře Median neprošly časové snímky v první vlně verifikací opakovaným vstupem, od druhé vlně bylo opakovaným pořízením kontrolováno 100 % diářů.

15.2. Vyřazování dotazníků

U CAPI individuálního a proxy dotazníku, PAPI dotazníků a časových diářů bylo zjišťováno dodržení věkové populace respondentů. Instrumenty vyplněné respondentem jiné věkové kategorie byly vyřazeny. Pokud se věk uvedený v PAPI dotazníku (*papiage*) a CAPI domácnostním dotazníku (*age*) lišil nejvýše o 1 rok a zároveň vyplněný instrument odpovídal věku z PAPI dotazníku, byla proměnná *age* upravena podle hodnoty proměnné *papiage*.

Z datového souboru byly ve druhém až čtvrtém roce výzkumu vyřazeny jednotky PAPI dotazníků a časových diářů, jejichž identifikační údaje odpovídaly bývalým členům domácnosti.

Ze zpracování byly vyřazeny PAPI dotazníky a časové snímky, které kvůli chybějícím identifikačním kódům nemohly být propojeny s ostatními dotazníky. V rámci jedné vlny se jednalo o jednotky až desítky papírových dotazníků.

Ve čtvrtém ročníku byl zajišťován písemný souhlas rodičů s vyplňováním dotazníků pro děti. Ze zpracování byly vyřazeny jednotky PAPI dotazníků a časových snímků, k nimž souhlas rodičů chyběl.

15.3. Tvorba datového souboru

Realizátoři předávali výzkumnému týmu oddělené datové soubory s proměnnými z jednotlivých instrumentů (CAPI dotazník pro domácnost byl v první vlně dále rozdělen na dva soubory s proměnnými na úrovni domácnosti a na úrovni jednotlivce). Soubory z jednotlivých instrumentů byly posléze sloučeny do jednoho datového souboru. Pokud obsahovalo více instrumentů tutéž otázku (např. CAPI individuální dotazník a dětské PAPI dotazníky), je v datovém souboru zařazena jako jedna proměnná (více o struktuře datového souboru v oddílu 5. *Obsah a struktura datového souboru*).

15.4. Čištění dat

U vztahů v domácnosti (proměnné *rel1* až *rel12*) byla zjišťována konzistence hodnot uvedených u jednotlivých členů domácnosti. Nekonzistentní nebo jinak nelogické vztahy (např. otec mladší než syn) byly zkontrolovány a opraveny podle nahrávek z dotazování nebo podle logiky ostatních vztahů a demografických proměnných.

Proměnné *name*, *sex*, *birthm*, *birthy* a částečně také *rel1* až *rel12* (neměnné vztahy) byly začištěny podle respondentů opravených hodnot z následujících ročníků výzkumů (CAPI dotazník pro domácnost nabízel respondentům od druhé vlny možnost opravy údajů zapsaných v předchozím roce) a došlo k příslušným úpravám proměnných filtrovaných na jejich základě.

Proměnné *mstat*, *estat*, *educ*, *std*, *stdch* a *rel1* až *rel12o2* byly individuálně začištěny podle hodnot návazných proměnných v příslušné vlně (např. respondent uvedl v otevřené odpovědi proměnné *stype*, že ve skutečnosti nestuduje), hodnot těchto proměnných zadaných v následujících vlnách a hodnot návazných proměnných v následujících vlnách (např. *mstat* bylo čištěno podle hodnot *mchange*) s přihlédnutím k celkové logice jednotlivých odpovědí. V návaznosti na změnu výchozích socio-demografik bylo provedeno potřebné začištění proměnných filtrovaných na jejich základě.

V proměnných *ustarty(x)*, *uendy(x)*, *pstarty* a *lstarty* byly neplatné odpovědi (nevím a odmítnutí) převedeny na kód 999 996, pokud hodnota *estat* z minulého roku odpovídala současnému *estat*. Těmto případům byly odpovědi v proměnných *ustartm*, *uendm*, *pstartm* a *lstartm* opraveny na hodnotu 96.

V proměnných *mstat*, *birthy* a *age* byly kódem 98 / 999 998 nahrazeny vysoce nepravděpodobné odpovědi (narození před rokem 1900, rodinný stav registrované partnerství, pokud ostatní údaje naznačovaly záměnu registrovaného partnerství s nesezdaným soužitím).

V proměnných *svise* a *empl* byla hodnota 1 převedena na hodnotu 2, pokud v proměnné *subor*, resp. *svise* uvedl respondent nulovou odpověď.

V proměnných *shours*, *hours*, *adhours*, *choosehr* a *hswork* byly hodnoty vyšší než 140 nahrazeny kódem 999 998. V proměnné *height* byly tímto kódem nahrazeny hodnoty nižší než 100 a vyšší než 250, v proměnné *weight* hodnoty nižší než 40 a vyšší než 250.

Pokud byly filtrační proměnné *papipart* a *papijob* bez odpovědi, ale v návazných otázkách respondent uvedl platné odpovědi, byla proměnná *papipart* opravena na hodnotu 1, pokud *part* = 1, a proměnná *papijob* na 1, pokud *estat* = 1/2/3/4/5/6/9 or *job* = 1.

Hodnoty proměnné *asame* byly srovnány s adresami domácností⁹ ze všech dosavadních ročníků výzkumu a v případné nesrovnalosti (např. stejná adresa v obou příslušných letech, ale proměnná *asame* značí změnu bydliště) byly vyřešeny za pomoci konzultace s tazateli. Konzultace s tazateli proběhly rovněž ve všech případech změny adresy podle proměnné *asame*, aby bylo možné určit, zda změna adresy je skutečnou změnou bydliště domácnosti, nebo zda adresa evidovaná v databázi z předchozí vlny byla nepřesně či chybně uvedena a domácnost bydliště nezměnila. Ve druhém případě byla v proměnné *asame* převedena hodnota odpovědi na 1 (stejná adresa) a vyčištěny návazné proměnné.

Proměnné vztahující se k dennímu a spánkovému režimu z PAPI dotazníku pro dospělé ve čtvrtém ročníku (*dayreg* až *daylifm*) byly čištěny tak, aby byla zachována formální logika vyplnění otázek, tj. aby se nevalidní kódy 96, 98 a 99 vztahovaly zároveň na údaj o hodině i minutě. Odpovědi spadající mimo myslitelné rozmezí dnů, hodin či minut a hodnoty nevyhovující předepsanému formátu odpovědí byly začištěny postupem znázorněným v tabulce 39 (v posledním sloupci tabulky je uveden počet případů, u nichž byla začištěna alespoň jedna proměnná z dané skupiny). Hodnoty 24 byly v údajích o hodinách ponechány pouze u proměnných *dayliwh* a *daylifh*, které vyjadřující trvání, nikoli časový okamžik. V ostatních hodinových proměnných byly hodnoty 24 překódovány na 0.

⁹ Adresy domácností jsou vedeny v dotazovém prostředí mimo CAPI dotazníky. Od druhého ročníku mohli tazatelé do systému zanášet novou adresu domácnosti.

Tabulka 39: čištění proměnných dayreg až daylifm

proměnné	čištěné hodnoty	výsledné hodnoty	případy
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	přepoččet na hodiny	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	přepoččet na hodiny	1
dayliwh, daylifh	H > 24	H = 98, M = 98	65
dayliwm, daylifm	M = 99, H <= 24	M = 0	119
dayliwm, daylifm	M >= 60	přepoččet na hodiny	15
dayliwh, daylifh	H = 99	H = 0	6

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

**) H = proměnná s údajem o hodině, M = proměnná s údajem o minutě

15.5. Kódování otevřených a polootevřených otázek

V polootevřených a otevřených otázkách *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* a *cpartywv* byly volné odpovědi zakódovány do širších kategorií v rámci textové proměnné nebo přiřazeny k odpovídající uzavřené možnosti. Nové kategorie v rámci textových proměnných vzniklé rekódováním volných výpovědí nejsou nutně vzájemně vylučné a vyčerpávající, protože do značné míry respektují obsah a míru podrobnosti původní výpovědi.

Volné odpovědi v otázkách na důvod přestěhování z předchozího bydliště (*rework, refam, rehous, reelse*) byly v relevantních případech přiřazeny k existujícím možnostem v rámci jiné skupiny důvodů přestěhování.

Proměnné s popisem pracovní činnosti (*swrkps, nwrkps, wrkps, pwrkps, hwrkps(x), fwrkps, mwrkps, jobas, jobex*) jsou zakódovány podle klasifikace CZ-ISCO.

15.6. Imputace hodnot

V datech z druhé až čtvrté vlny byly u proměnných *htnr, isco08, svise, subor, empl, empls, fsize* a *contr* z CAPI individuálního dotazníku doplněny hodnoty z předchozího ročníku, pokud respondent potvrdil při podmíněném dotazování jejich platnost (více informací v oddílu 8. *Proměnné v podmíněném dotazování*).

V PAPI dotaznících neproběhla žádná imputace dat.

V časových diářích byl do odpovídajících časových bloků doplněn kód 1 „spánek“, pokud záznam respondenta skončil ve večerních hodinách a dále nepokračoval nebo naopak začal až v druhém či dalším bloku v ranních hodinách. Imputace je v datech z časových snímků značena. Podrobnosti jsou obsaženy v samostatné dokumentaci k datům z časových snímků.

15.7. Úpravy počtu případů podle navazujících vln

Po ukončení sběru dat ve druhém roce výzkumu bylo do dat z první vlny doplněno 9 osob, které žily v dotázaných domácnostech již v prvním roce, ale nebyly tazatelem zaznamenány. Zároveň bylo z těchto dat vyřazeno 41 domácností, u nichž se během druhé vlny odhalilo závažná pochybení v realizaci rozhovorů v prvním roce (započítáno v *tabulce 3: výsledky kontaktů ve vzorku domácností, vlna 1*, jako výsledek „dotázáno, vyřazeno na základě kontrol“).

V návaznosti na dotazování ve třetím roce bylo do dat z první vlny doplněno dalších 29 dříve opomenutých osob a do dat z druhé vlny 15 osob. Ke zpětnému vyřazování domácností kvůli pochybení tazatelů již po třetí vlně nedošlo.

Na základě sběru dat ve čtvrtém ročníku bylo do dat z první vlny přidáno 6 osob, do dat z druhé vlny 7 osob a do dat ze třetí vlny 8 osob.

15.8. Neplatné odpovědi/hodnoty

Všechny otázky v CAPI dotaznících (s výjimkou kognitivních testů v individuálním CAPI dotazníku z druhé vlny) dovozovaly tazateli zaznamenat situaci, kdy respondent nebyl schopen nebo odmítl na otázku odpovědět. Při osobním rozhovoru tazatel respondentům tyto možnosti přímo nepředčítal a nebyly vypsány na kartách, v CAWI mutaci byly uvedeny za platnými možnostmi. V publikovaných CAPI dotaznících jsou nenabízené možnosti uvedeny v hranatých závorkách.

V PAPI dotaznících a jejich CAWI mutaci povolovaly některé otázky explicitně možnost „nevím“. Ve vybraných otázkách (např. politické znalosti *knowkal* až *knowdsss*) však není odpověď „nevím“ značena kódem neplatné hodnoty (88).

Při pořizování a čištění datového souboru byly do proměnných z CAPI a PAPI doplňovány chybové kódy 98 / 999 998 v případech popsaných v *tabulce 40*.

Kódy neplatných hodnot jsou v SPSS datových souborech nastaveny jako user-missing. Jako user-missing jsou dále nastaveny kódy 96 / 999 996 v proměnných *sentryy*, *sentrym*, *ustartm(x)*, *ustarty(x)*, *uendm(x)*, *uendy(x)*, *pstartm*, *pstarty*, *dcmoney* až *dcsave*, *cless(x)*, *relparnl*, *fowny*, *fcoopy*, *hendy(x)*, *mleavee(x)*, *uneme(x)*, *nseentryy(x)*, *nsetrym(x)*, *bedwh*, *bedwm*, *sleepwh*, *sleepwm*, *wakewh*, *wakewm*, *bedfh*, *bedfm*, *sleepfh*, *sleepfm*, *wakefh*, *wakefm*, *symods* až *symssso*. Ačkoli se jedná o respondentům nabízené platné možnosti, značí odpověď, která se nachází mimo hlavní škálu variant odpovědí. Jako user-missing nejsou nastaveny odpovědi „nevím“ u PAPI proměnných, pokud nemají kód neplatné hodnoty (88).

Tabulka 40: neplatné odpovědi a jejich kódy v CAPI a PAPI datech

88 / 888 888	CAPI	nevím, nemohu posoudit – nebylo nabízeno <i>V CAWI mutaci dotazníku uvedeno mezi možnostmi.</i>
	PAPI	nevím (kódem neplatné odpovědi značeno jen u vybraných proměnných)
99 / 999 999	CAPI	odmítl(a) – nebylo nabízeno <i>V CAWI mutaci dotazníku uvedeno mezi možnostmi ve znění „Nechci uvést“.</i>
	PAPI	bez odpovědi
97 / 999 997	CAPI/PAPI	hodnota pod filtrem
98 / 999 998	CAPI	otázky nebyly dotázány z důvodu navolení chybných filtračních hodnot (které byly opraveny během čištění)
	CAPI	<i>etran</i> , <i>ohousch</i> , <i>nown</i> , <i>part</i> , <i>sumchild</i> , <i>nage(x)</i> <i>dignum</i> , <i>c_dignum</i> , <i>dinum</i> , <i>c_dinum</i> , <i>digcam</i> , <i>dincam</i> , <i>signum</i> , <i>sinum</i> , <i>penum</i> , <i>benum</i> , <i>cognum</i> = nelze odvodit hodnotu generované proměnné, protože některá z výchozích proměnných obsahuje neplatnou odpověď
	CAPI	<i>birthy</i> , <i>age</i> , <i>mstat</i> = chybná odpověď (narození před rokem 1900, rodinný stav registrované partnerství, pokud ostatní údaje naznačovaly záměnu registrovaného partnerství s nesezdaným soužitím)
	CAPI/PAPI	<i>shours</i> , <i>hours</i> , <i>adhours</i> , <i>choosehr</i> , <i>weight</i> , <i>height</i> , <i>hswork</i> = odpověď spadající mimo rozsah přípustných hodnot
	CAPI/ PAPI	<i>isco08</i> , <i>fisco08</i> , <i>misco08</i> , <i>aisco08</i> , <i>eisco08</i> (včetně odvozených ukazatelů ISEI a ESeC) = výchozí otevřená odpověď nelze zakódovat
	PAPI	nejednoznačná odpověď (např. výběr dvou odpovědí bez značení, která platí; nečitelná odpověď v otevřené otázce)

V prvním a druhém ročníku výzkumu došlo v dotazových skriptech CAPI domácnostního a individuálního dotazníku k několika technickým problémům, v jejichž důsledku nebyly u všech relevantních respondentů dotázány některé otázky nebo hodnoty dotázaných otázek nebyly uloženy. Ve druhé vlně byly problémy opraveny v průběhu terénních prací a chybějící hodnoty byly dodatečně zjišťovány telefonicky v průběhu září a října 2016.

Chybějící hodnoty vzniklé tímto způsobem jsou v datech značeny kódy 98 / 999 998.

Tabulka 41: CAPI proměnné s chybějícími hodnotami v důsledku problémů scriptu		W1	W2
estat	na jednom typu tazatelských laptopů byly kategorie 8, 9, 10 a 11 viditelné jen po horizontálním posunu obrazovky a zvýšily se tím frekvence viditelných kategorií 12 a 14. Změny ekonomického postavení z kategorií 8, 9 a 11 na 12 a 14 byly opraveny zpět na hodnoty 8, 9 a 11.	-	1
exutil, exrent, exmort, exrep	chybný výpočet filtrační proměnné <i>ohousch</i>	-	1
chben	chybný výpočet filtrační proměnné <i>p0to26</i> u části relevantních domácností	1	-
emplch	chybná implementace filtru (vynechání podmínky <i>w1_empl = 2</i>)	-	1
svise, subor, fsize	část hodnot nebyla uložena – chyba v komunikaci dotazníku a FTP u <i>agen = 0</i>	1	-
jbsat	chybná implementace filtru (vynechání podmínky <i>estat = 4</i>)	-	1
inval	chybná implementace filtru (vynechání podmínky <i>inot = 1</i> a <i>inhous = 1</i>)	-	1
rplaidh(x)	chybná implementace filtru (vynechání podmínky <i>raidh = 97</i>)	-	1
nplaidh(x)	chybná implementace filtru (vynechání podmínky <i>naidh = 97</i>)	-	1

16. Specifika vzorku

V datech CHPS jsou ve srovnání s údaji ze Sčítání lidu, domů a bytů 2011¹⁰ v menší míře zastoupeny jednočlenné domácnosti (22 % vs. 33 %). Podhodnocení jednočlenných domácností může být dáno nižší šancí tyto domácnosti zastihnout (menší pravděpodobnost, že jediný člen bude doma, ve srovnání s pravděpodobností u vícečlenných domácností, že kterýkoli dospělý člen bude doma), jejich menší ochotou k účasti (např. obavy seniorů v jednočlenných domácnostech z tazatele), systémem odměňování tazatelů za jednotlivé dotazové materiály, který vede k preferenci získávání větších domácností, nebo nedostatečným rozdělováním bytových domácností na hospodařící.

Kompletní soubor všech členů domácnosti lze rozdělit na dílčí skupiny podle vyplnění jednotlivých typů instrumentů (např. soubor respondentů CAPI individuálního dotazníku). Struktura těchto dílčích skupin se liší jak mezi sebou, tak zejména ve srovnání s kompletním souborem. Dochází zde k vychýlení v důsledku toho, že respondenti nebyli vybíráni náhodně, ale cílem bylo dotázat všechny členy domácnosti. Ve spojení s neúplnou návratností v rámci domácnosti vede tato skutečnost k vyššímu zastoupení dostupnějších sociodemografických segmentů populace (ženy, starší lidé atp.) ve srovnání se souborem všech členů domácnosti.

¹⁰ Sčítání lidu, domů a bytů 2011, Český statistický úřad. Výsledky na: <https://www.czso.cz/csu/sldb/domacnosti>

17. Post-stratifikační váhy

17.1. Binární proměnné určující vyplnění instrumentů

Datový soubor obsahuje post-stratifikační váhy, které umožňují korigovat odchylky od populačních proporcí vzhledem k pohlaví, věku, vzdělání a kraji bydliště. Pro respondenty časových snímků jsou navíc k dispozici váhy vyrovnávající zastoupení dnů v týdnu.

Protože jednotlivé instrumenty a jejich kombinace vyplnila vždy jiná skupina respondentů, bylo připraveno 16 proměnných s vahami vztahujícími se k různým instrumentům a jejich kombinacím. Pro 6 dílčích souborů respondentů, kteří vyplnili časový snímek, byly váhy vypočítány ve dvou variantách (první varianta váží pouze na základě sociodemografických proměnných, druhá varianta navíc vedle korekce sociodemografik zajišťuje uniformní distribuci dnů v týdnu). Celkově se tedy v datech nachází 22 vážících proměnných.

Pro výběr jednotlivých skupin respondentů slouží binární proměnné *hh_hh* až *child_papi_diary*, které určují, zda člen domácnosti vyplnil daný instrument (bez ohledu na to, zda patřil do populace instrumentu). Pokud ano (1), je mu přiřazena hodnota vážící proměnné pro daný instrument nebo kombinaci instrumentů. Přehled binárních indikátorů a příslušných vah obsahuje *tabulka 42*, konstrukci dílčích vzorků respondentů ilustruje *tabulka 45*.

V datovém souboru z druhého ročníku mají bývalí členové domácnosti (*rstat* = 5) v uvedených binárních proměnných přidělenou hodnotu 0 a hodnota vážící proměnné pro ně není spočtena.

Tabulka 42: indikátory skupin respondentů a příslušné váhy

hh_hh	člen domácnosti, který vyplnil CAPI dotazník pro domácnost
W_hh_hh	váha pro domácnosti
hh_indi	člen dotázané domácnosti (tj. všichni členové kromě bývalých)
W_hh_indi	váha pro členy domácností
indi	člen domácnosti, který vyplnil CAPI individuální dotazník
W_indi	váha pro respondenty CAPI individuálního dotazníku
indi_proxy	člen domácnosti, který vyplnil CAPI individuální dotazník, NEBO byl za něj vyplněn CAPI proxy dotazník
W_indi_proxy	váha pro respondenty CAPI individuálního dotazníku NEBO respondenty, za něž byl vyplněn CAPI proxy dotazník
papi_adult	člen domácnosti, který vyplnil PAPI dotazník pro dospělé
W_papi_adult	váha pro respondenty PAPI dotazníku pro dospělé
papi_child	člen domácnosti, který vyplnil PAPI dotazník pro děti
W_papi_child	váha pro respondenty PAPI dotazníků pro děti
adult_child	člen domácnosti, který vyplnil PAPI dotazník pro dospělé NEBO PAPI dotazník pro děti

W_adult_child	váha pro respondenty PAPI dotazníku pro dospělé NEBO PAPI dotazníků pro děti
adult_child15	člen domácnosti, který vyplnil PAPI dotazník pro dospělé NEBO PAPI dotazník pro děti 15–17 let
W_adult_child15	váha pro respondenty PAPI dotazníku pro dospělé NEBO PAPI dotazníku pro děti 15–17 let
indi_child15	člen domácnosti, který vyplnil CAPI individuální dotazník NEBO PAPI dotazník pro děti 15–17 let
W_indi_child15	váha pro respondenty CAPI individuálního dotazníku NEBO PAPI dotazníku pro děti 15–17 let
diary_adult	člen domácnosti, který vyplnil časový snímek pro dospělé
W_diary_adult	váha pro respondenty časového snímku pro dospělé
Wd_diary_adult	váha pro respondenty časového snímku pro dospělé s uniformní distribucí dnů
diary_child	člen domácnosti, který vyplnil časový snímek pro děti
W_diary_child	váha pro respondenty časového snímku pro děti
Wd_diary_child	váha pro respondenty časového snímku pro děti zohledňující distribuci dnů v týdnu
indi_papi	člen domácnosti, který vyplnil CAPI individuální dotazník A ZÁROVEŇ PAPI dotazník pro dospělé
W_indi_papi	váha pro respondenty CAPI individuálního dotazníku A ZÁROVEŇ PAPI dotazníku pro dospělé
indi_diary	člen domácnosti, který vyplnil CAPI individuální dotazník A ZÁROVEŇ časový snímek pro dospělé
W_indi_diary	váha pro respondenty CAPI individuálního dotazníku A ZÁROVEŇ časového snímku pro dospělé
Wd_indi_diary	váha pro respondenty CAPI individuálního dotazníku A ZÁROVEŇ časového snímku pro dospělé s uniformní distribucí dnů
adult_papi_diary	člen domácnosti, který vyplnil PAPI dotazník pro dospělé A ZÁROVEŇ časový snímek pro dospělé
W_adult_papi_diary	váha pro respondenty PAPI dotazníku pro dospělé A ZÁROVEŇ časového snímku pro dospělé
Wd_adult_papi_diary	váha pro respondenty PAPI dotazníku pro dospělé A ZÁROVEŇ časového snímku pro dospělé s uniformní distribucí dnů
indi_papi_diary	člen domácnosti, který vyplnil CAPI individuální dotazník A ZÁROVEŇ PAPI dotazník pro dospělé A ZÁROVEŇ časový snímek pro dospělé
W_indi_papi_diary	váha pro respondenty CAPI individuálního dotazníku A ZÁROVEŇ PAPI dotazníku pro dospělé A ZÁROVEŇ časového snímku pro dospělé
Wd_indi_papi_diary	váha pro respondenty CAPI individuálního dotazníku A ZÁROVEŇ PAPI dotazníku pro dospělé A ZÁROVEŇ časového snímku pro dospělé s uniformní distribucí dnů
child_papi_diary	člen domácnosti, který vyplnil PAPI dotazník pro děti A ZÁROVEŇ časový snímek pro děti
W_child_papi_diary	váha pro respondenty PAPI dotazníku pro děti A ZÁROVEŇ časového snímku pro děti
Wd_child_papi_diary	váha pro respondenty PAPI dotazníku pro děti A ZÁROVEŇ časového snímku pro děti s uniformní distribucí dnů

17.2. Metoda vážení

Vážení bylo provedeno v programu R funkcí *rake* (*survey* package), která pomocí iterativní procedury vyrovnává vážené marginální distribuce vzorku a známé populační distribuce. Maximální velikost váhy byla zafixována na hodnotě 4 (funkce *trim*). V datovém souboru z 1. vlny byla fixace využita u vah *Wd_diary_adult*, *Wd_indi_diary*, *Wd_adult_papi_diary*, *Wd_indi_papi_diary*, v datovém souboru z 2. vlny u vah *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*, v datovém souboru z 3. vlny u vah *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* a v datovém souboru ze 4. vlny u vah *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*.

R script pro provedení vážení je dostupný jako samostatný soubor.

17.3. Proměnné pro vážení

V závislosti na instrumentu byly využity tři skupiny proměnných pro vážení:

- kraj bydliště (NUTS 3),
- cross-klasifikace pohlaví, vzdělání a věku + kraj bydliště (NUTS 3),
- pohlaví + kraj bydliště (NUTS 3).

Zdrojem kontrolních distribucí pro vážení byla data z Výběrového šetření pracovních sil z příslušného roku (2015 pro první ročník CHPS, 2016 pro druhý ročník atd.), Sčítání lidu, domů a bytů (SLBD) 2011 a údaje Českého statistického úřadu o věkovém složení obyvatel k 1. červenci příslušného roku.

Tabulka 43: skupiny proměnných pro vážení

1	kraj bydliště (NUTS 3)	kraj: SLDB 2011 (počet hospodařících domácností podle krajů)
2	cross-klasifikace pohlaví, vzdělání a věku + kraj bydliště (NUTS 3)	kraj: Věkové složení obyvatel k 1. 7. příslušného roku (ČSÚ) pohlaví, věk, vzdělání: VŠPS z příslušného roku ¹¹
3	pohlaví + kraj bydliště (NUTS 3)	kraj, pohlaví: Věkové složení mužů k 1. 7. příslušného roku a Věkové složení žen k 1. 7. příslušného roku (ČSÚ)

Proměnné *sex* (pohlaví) a *region* (kraj) nebyly pro vážení upravovány. Kontinuální proměnná *age* byla v závislosti na populaci daného instrumentu rozdělena do pěti, případně čtyř věkových kategorií (celkem 3 varianty kategorizace věku). Proměnná *educ* (vzdělání) byla rozdělena do 4 kategorií (tabulka 44). Kontrolní data využitá pro odvození teoretických distribucí jsou dostupná jako samostatný MS Excel soubor.

¹¹ Nevážený roční soubor.

Váhy pro respondenty časových snímků byly spočteny ve dvou variantách. První varianta pracuje pouze s uvedenými sociodemografickými proměnnými, druhá varianta zároveň s úpravou sociodemografik převažuje zastoupení dnů v týdnu na uniformní distribuci. Uniformní distribuce dnů v týdnu byla do vážení zařazena jako samostatné kritérium, a nikoli v křížení s dalšími proměnnými.

Tabulka 44: kategorie proměnných pro vážení

kraj (<i>region</i>)	1–14	dle hodnot proměnné <i>region</i>						
pohlaví (<i>sex</i>)	1	muž (<i>sex</i> = 1)						
	2	žena (<i>sex</i> = 2)						
věk1 (<i>ageR</i>)	0	chybějící h.	0	chybějící h.	0	chybějící h.		
	1	$age \leq 17$	věk2 (<i>ageR2</i>)	1	$10 \geq age \leq 34$	věk3 (<i>ageR3</i>)	1	$15 \geq age \leq 34$
	2	$18 \geq age \leq 34$	2	$35 \geq age \leq 54$	2	$35 \geq age \leq 54$		
	3	$35 \geq age \leq 54$	3	$age \geq 55$	3	$age \geq 55$		
	4	$age \geq 55$						
vzdělání (<i>educR</i>)	0	chybějící hodnota						
	1	bez vzdělání, základní vzdělání (<i>educ</i> = 0, <i>educ</i> = 1, <i>educ</i> = 2)						
	2	střední vzdělání (<i>educ</i> = 3), střední vzdělání s výučním listem (<i>educ</i> = 4), střední vzdělání s maturitou (<i>educ</i> = 5, <i>educ</i> = 6, <i>educ</i> = 7)						
	3	vyšší odborné (<i>educ</i> = 8) a vysokoškolské vzdělání (<i>educ</i> = 9, <i>educ</i> = 10, <i>educ</i> = 11)						

17.4. Práce s chybějícími hodnotami

Jak data CHPS, tak VŠPS obsahují chybějící hodnoty v proměnných věk a vzdělání. Práce s chybějícími hodnotami se řídila následujícími pravidly:

- 1 Chybějící hodnoty v kontrolních datech i datech ze vzorku: Buňky s chybějícími hodnotami jsou využity standardním způsobem (v žádném z případů nebyl počet chybějících odpovědí ve vzorku větší než odpovídající počet chybějících odpovědí v kontrolních datech).
- 2 Chybějící hodnoty jen v datech ze vzorku: hodnoty v polích s chybějícími odpověďmi jsou vykopírovány do odpovídajících polí kontrolní tabulky (za předpokladu MAR). Zbylé hodnoty v kontrolní tabulce jsou proporčně upraveny tak, aby celkový počet případů a poměr mezi hodnotami v polích tabulky zůstanou nezměněny.
- 3 Chybějící hodnoty jen v kontrolních datech: chybějící hodnoty v kontrolní tabulce jsou zanedbány (za předpokladu MCAR). Podíl chybějících hodnot v kontrolních datech nepřesahoval 0,01 %.

Tabulka 45: přehled dílčích vzorků respondentů a jejich vážení

	CAPI dom.	CAPI indi.	CAPI prox.	PAPI dosp.	PAPI děti 15–17	PAPI děti 10–14	čas. sním. dosp.	čas. sním. děti	n (vlna 1)	n (vlna 2)	n (vlna 3)	n (vlna 4)	proměnné vážení	kat. věku
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	&				+	+			866	602	644	572	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 případy, kdy byl vyplněn takto značený instrument

+ případy, kde byl vyplněn alespoň jeden z takto značených instrumentů

& případy, kdy byly vyplněny všechny takto značené instrumenty

18. Verze datového souboru

Datový soubor z prvního ročníku CHPS byl ke dni 18. 2. 2019 vydán v pěti verzích.

Tabulka 46: verze datového souboru z 1. ročníku

	datum zveřejnění	popis změn
v1.0	6. 2. 2017	-
v1.1	6. 3. 2017	vynechání proměnných <i>read</i> , <i>parnec</i> , <i>gchrep</i> , <i>gchclean</i> , <i>gchmeal</i> , <i>gchlaund</i>
v2.0	11. 4. 2017	<ul style="list-style-type: none">úprava popisek vybraných proměnných v souladu s labely používanými v datovém souboru z druhého ročníku CHPS<i>jobtp</i> (<i>agen</i> = 0): oprava chybného čištění dat (převod hodnot 99 na 1/2/88)<i>svise</i>: oprava chybného čištění dat (převod hodnot 97 na 1/2/88 v souvislosti s opravou proměnné <i>jobtp</i>); doplnění chybových kódů 98<i>subor</i> (<i>agen</i> = 0): oprava chybného čištění dat (převod hodnot 999 997 na 1–1 000/888 888 v souvislosti s opravou proměnné <i>jobtp</i>); doplnění chybových kódů 999 998<i>empl</i> (<i>agen</i> = 0): oprava chybného čištění dat (převod hodnot 97 na 1/2 v souvislosti s opravou proměnné <i>jobtp</i>)<i>empls</i> (<i>agen</i> = 0): oprava chybného čištění dat (převod hodnot 999 997 na 1–15 v souvislosti s opravou proměnné <i>jobtp</i>)<i>fsize</i> (<i>agen</i> = 0): oprava chybného čištění dat (převod hodnot 97 na 1–6/88 v souvislosti s opravou proměnné <i>jobtp</i>); doplnění chybových kódů 98<i>fsubor</i> (<i>agen</i> = 0): oprava chybného čištění dat (převod hodnot 99 na 1/2/88)<i>ESeC08</i>: změna hodnoty v souvislosti s opravou proměnné <i>jobtp</i><i>ESeC08_6</i>: změna hodnoty v souvislosti s opravou proměnné <i>jobtp</i><i>read</i> (<i>agen</i> = 0, <i>papi</i> = 3): oprava chybných hodnot způsobených špatným nastavením pořizovací masky při scanování<i>parnec</i> (<i>agen</i> = 0, <i>papi</i> = 2): oprava chybných hodnot způsobených špatným nastavením pořizovací masky při scanování<i>gchrep</i> (<i>agen</i> = 0, <i>papi</i> = 3): oprava chybných hodnot způsobených špatným nastavením pořizovací masky při scanování<i>gchclean</i> (<i>agen</i> = 0, <i>papi</i> = 3): oprava chybných hodnot způsobených špatným nastavením pořizovací masky při scanování<i>gchmeal</i> (<i>agen</i> = 0, <i>papi</i> = 3): oprava chybných hodnot způsobených špatným nastavením pořizovací masky při scanování<i>gchlaund</i> (<i>agen</i> = 0, <i>papi</i> = 3): oprava chybných hodnot způsobených špatným nastavením pořizovací masky při scanováníproměnné z PAPI dotazníků (<i>agen</i> = 0): doplnění hodnot nepřechtených při scanování (převod hodnot 99 na platné)<i>exelseo</i>, <i>cpartywv</i>: úprava kódovacího schématu v souladu s kódováním užívaným v datovém souboru z druhého ročníku CHPS

v3.0	6. 2. 2018	<ul style="list-style-type: none"> • přejmenování identifikátorů <i>hid</i> a <i>pid</i> na <i>w1_hid</i> a <i>w1_pid</i> • úprava sociodemografických proměnných na základě údajů z 3. ročníku • doplnění 29 členů domácnosti, kteří byli v 1. ročníku opomenuti • úprava popisků vybraných proměnných • úprava nastavení šířky vybraných stringových proměnných pro harmonizaci s datovým souborem z vlny 2 • úprava výpočtu proměnné <i>part</i> (zahrnuto doplnění podle hodnoty <i>papipart</i>) • oprava chyby ve výpočtu proměnné <i>sumchild</i> • proměnná <i>pstartm</i> (měsíc začátku CAPI dotazování) přejmenována na <i>pstartmo</i> z důvodu shody původního názvu s kódem nové proměnné ve 2. ročníku výzkumu • přidání proměnné <i>intid</i> s kódem tazatele • přidání proměnných <i>dignum – cognum</i> • přidání proměnných <i>Wd_diary_adult</i>, <i>Wd_diary_child</i>, <i>Wd_indi_diary</i>, <i>Wd_adult_papi_diary</i>, <i>Wd_indi_papi_diary</i>, <i>Wd_child_papi_diary</i>
v4.0	18. 2. 2019	<ul style="list-style-type: none"> • úprava sociodemografických proměnných na základě údajů ze 4. ročníku • doplnění 6 členů domácnosti, kteří byli v 1. ročníku opomenuti • přepočty vah v návaznosti na přidání nových členů a opravu sociodemografických proměnných • doplnění dosud chybějících popisků hodnot relevantních pro některou z ostatních vln • sjednocení formátů proměnných mezi vlnami • úprava popisků vybraných proměnných a hodnot • úprava nastavení user-missing hodnot (převedení diskrétních hodnot na rozsahy) • oprava čištění filtrů (zejména <i>maben</i>, <i>maliv</i>, <i>faliv</i>) • drobné opravy kódování stringových proměnných (např. <i>cpartywv</i>) • <i>nage1 – nage8</i>: doplnění proměnných s věky nerezidenčních dětí • <i>dinum</i>: oprava chyby ve výpočtu proměnné • <i>h_papi_child</i>, <i>h_diary_child</i>: nahrazení nulových hodnot system missingy v domácnostech bez dětí ve věku 10 až 17 let

Datový soubor z druhého ročníku CHPS byl ke dni 18. 2. 2019 vydán ve dvou verzích.

Tabulka 47: verze datového souboru z 2. ročníku

	datum zveřejnění	popis změn
v1.0	6. 2. 2018	-
v2.0	18. 2. 2019	<ul style="list-style-type: none"> • úprava sociodemografických proměnných na základě údajů ze 4. ročníku • doplnění 7 členů domácnosti, kteří byli v 2. ročníku opomenuti • přepočty vah v návaznosti na přidání nových členů a opravu sociodemografických proměnných • doplnění dosud chybějících popisků hodnot relevantních pro některou z ostatních vln • sjednocení formátů proměnných mezi vlnami • úprava popisků vybraných proměnných a hodnot

- úprava nastavení user-missing hodnot (převedení diskrétních hodnot na rozsahy)
- oprava čištění filtrů (zejména *eduex2*, *breakup*)
- drobné opravy kódování stringových proměnných (např. *cpartywv*)
- *hendreo7* – *hendreo10*: úprava nastavení šířky stringových proměnných
- *partycl*, *partyclo*, *partywv*, *partywvo*: zakódování stran uvedených ve volné odpovědi do kódů stran nabízených jako varianty odpovědi od 3. ročníku
- přidání proměnných nerezidenčních dětí nevyužitého pořadí (6 až 8)
- *nage1* – *nage8*: oprava chyby ve výpočtu proměnné
- *h_papi_child*, *hp_papi_child*, *h_diary_child*, *hp_diary_child*: nahrazení nulových hodnot system missingy v domácnostech bez dětí ve věku 10 až 17 let
- *h_indi*, *hp_indi*, *h_indi_proxy*, *hp_indi_proxy*, *h_papi_adult*, *hp_papi_adult*, *h_diary_adult*, *hp_diary_adult*: nahrazení nulových hodnot hodnotou s počtem/podílem dotazníků v domácnosti u případů *rstat* = 5

Datový soubor z třetího ročníku CHPS byl ke dni 18. 2. 2019 vydán v jedné verzi.

Tabulka 48: verze datového souboru z 3. ročníku

	datum zveřejnění	popis změn
v1.0	18. 2. 2019	-

Datový soubor ze čtvrtého ročníku CHPS byl ke dni 21. 11. 2019 vydán v jedné verzi.

Tabulka 49: verze datového souboru z 3. ročníku

	datum zveřejnění	popis změn
v1.0	21. 11. 2019	-

19. Přehled materiálů k 1. až 4. ročníku CHPS

Tabulka 50: přehled materiálů k 1. až 4. ročníku CHPS		W1	W2	W3	W4
Kontaktní materiály					
Avizní dopis pro respondenty	varianta MEDIAN a SM	1	1	1	1
Informační leták pro respondenty	varianta MEDIAN a SM	1	1	1	1
Dopis obcím	varianta MEDIAN a SM	1	-	-	-
Dopis městům		1	-	-	-
Dopis krajským ředitelstvím Policie ČR		1	-	-	-
Navštívenka	varianta MEDIAN a SM	1	1	1	1
Děkovka	varianta MEDIAN a SM	1	1	1	1
Dotazové instrumenty					
CAPI dotazník pro domácnost		1	1	1	1
CAPI individuální dotazník		1	1	1	1
CAPI proxy dotazník		1	1	1	1
PAPI dotazník pro dospělé *)	varianta MEDIAN a SM	1	1	1	1
PAPI dotazník pro děti 15-17 let *)	varianta MEDIAN a SM	1	1	-	-
PAPI dotazník pro děti 10-14 let *)	varianta MEDIAN a SM	1	1	-	-
PAPI dotazník pro děti 10-17 let *)	varianta MEDIAN a SM	-	-	1	1
Časový snímek pro dospělé *)		1	1	1	1
Časový snímek pro děti *)		1	1	1	1
Karty ke CAPI dotazníku pro domácnost		1	1	1	1
Karty ke CAPI individuálnímu dotazníku		1	1	1	1
Karty ke CAPI proxy dotazníku		1	1	1	1
Záznamový list pro test slovní plynulosti		-	1	-	-
Časová osa pro záznam zaměstnanecké historie		-	1	-	-
Datový soubor					
Datové soubory z CAPI a PAPI dotazníků		1	1	1	1
Datové soubory z časových snímků pro děti a dospělé		1	1	1	1
Datové soubory z kontaktních formulářů		1	1	1	1
Dokumentace a další materiály					
Dokumentace k souborům z CAPI a PAPI dotazníků		1	1	1	1
Seznam proměnných v souboru z CAPI a PAPI dotazníků		1	1	1	1
SPSS syntax pro konstrukci ukazatelů ISEI a ESeC	v anglickém jazyce	1	- **)	- **)	- **)
SPSS syntax pro konstrukci příjmových proměnných	v anglickém jazyce	1	- **)	- **)	- **)
R script pro provedení vážení	v anglickém jazyce	1	1	1	1
Kontrolní distribuce využitě při vážení	v anglickém jazyce	1	1	1	1
Dokumentace k datům z časových snímků		1	1	1	1
Dokumentace k datům z kontaktních formulářů		1	1	1	1

*) Dotazník distribuovaný respondentům neobsahoval kódy jednotlivých proměnných.

***) Syntax pro druhý ročník je shodná se syntaxí z prvního ročníku.