

ČESKÉ PANELOVÉ ŠETŘENÍ DOMÁCNOSTÍ

Dokumentace k datům

Vlna 1, 2015

Mgr. Michaela Röschová

Sociologický ústav AV ČR, v.v.i.

www.promenyceskespolecnosti.cz

promeny.spolecnosti@soc.cas.cz

Verze 1.0 (27. 1. 2017)



Obsah

Shrnutí 4

1.	České panelové šetření domácností	5
2.	Výběrový soubor	7
2.1.	Cílová skupina	7
2.2.	Jednotka výběru.....	7
2.3.	Výběr vzorku domácností	8
3.	Sběr dat.....	10
3.1.	Školení tazatelů.....	10
3.2.	Pilotní šetření.....	10
3.3.	Avizování výzkumu.....	11
3.4.	Průběh terénních prací	11
3.5.	Kontaktování domácností	12
3.6.	Monitorování terénních prací	13
3.7.	Rozhovor v domácnosti	13
3.8.	Odměny pro respondenty a tazatele	14
3.9.	Kontrola práce tazatele.....	14
3.10.	Udržování kontaktu s domácnostmi	14
4.	Dotazové instrumenty	15
4.1.	Typy instrumentů.....	15
4.2.	Distribuce instrumentů	16
4.3.	Návratnost na úrovni instrumentů	16
4.4.	Dokumentace instrumentů.....	18
5.	Obsah a struktura datového souboru	20
5.1.	Řazení proměnných	20
5.2.	Struktura souboru.....	21
5.3.	Proměnné nezařazené do datového souboru.....	21
6.	Identifikační proměnné.....	23
7.	Generované proměnné	24
8.	Stratifikační postavení	28
9.	Proměnné vztahující se k průběhu a podmínkám dotazování	29
10.	Anonymizace dat	32
11.	Kontrola, zpracování a čištění dat.....	33

11.1.	Kontroly dat	33
11.2.	Vyřazování dotazníků	34
11.3.	Tvorba datového souboru.....	34
11.4.	Čištění dat	34
11.5.	Kódování otevřených a polootevřených otázek.....	35
11.6.	Imputace hodnot	35
11.7.	Úpravy počtu případů podle vlny 2	35
11.8.	Neplatné odpovědi/hodnoty.....	36
12.	Specifika vzorku.....	37
13.	Post-stratifikační váhy	38
13.1.	Binární proměnné určující vyplnění instrumentů	38
13.2.	Metoda vážení	39
13.3.	Proměnné pro vážení.....	40
13.4.	Práce s chybějícími hodnotami	41
14.	Přehled materiálů k prvnímu ročníku CHPS.....	43

Přehled tabulek

Tabulka 1: tematické okruhy Českého panelového šetření domácností (CHPS)	5
Tabulka 2: věkové vymezení populace jednotlivých instrumentů	7
Tabulka 3: výsledky kontaktů ve vzorku domácností	11
Tabulka 4: důvody odmítnutí účasti ve výzkumu	13
Tabulka 5: počty vyplněných instrumentů	17
Tabulka 6: míra vyplnění instrumentů celkem a podle typu domácnosti (hhtype)	17
Tabulka 7: pořadí skupin proměnných v datovém souboru	20
Tabulka 8: schéma struktury datového souboru	21
Tabulka 9: proměnné nezařazené do datového souboru	22
Tabulka 10: identifikační proměnné jednotlivce/domácnosti	23
Tabulka 11: identifikační čísla dalších členů domácnosti	23
Tabulka 12: generované proměnné na úrovni domácnosti	24
Tabulka 13: ukazatele přítomnosti skupin osob v domácnosti	25
Tabulka 14: typ domácnosti podle struktury	26
Tabulka 15: generované proměnné na úrovni jednotlivce	27
Tabulka 16: ukazatele stratifikačního postavení	28
Tabulka 17: místo bydliště domácnosti	29
Tabulka 18: charakteristiky dotazování	29
Tabulka 19: datum dotazování	30
Tabulka 20: Vyplnění dotazníků v domácnosti	31
Tabulka 21: neplatné odpovědi a jejich kódy v CAPI a PAPI datech	36
Tabulka 22: indikátory skupin respondentů a příslušné váhy	38
Tabulka 23: skupiny proměnných pro vážení	40
Tabulka 24: kategorie proměnných pro vážení	40
Tabulka 25: přehled dílčích vzorků respondentů a jejich vážení	42
Tabulka 26: přehled materiálů k prvnímu ročníku CHPS	43

Přehled schémat

Schéma 1: popis náležitostí publikovaných CAPI dotazníků	18
Schéma 2: srovnání původní a publikované podoby PAPI dotazníků	19

Shrnutí

Cílová skupina

Neinstitucionalizovaná populace České republiky

Jednotka výběru

Hospodařící domácnost

Předmět šetření

Všichni členové vybraných domácností

Výběr vzorku

Dvoustupňový stratifikovaný náhodný výběr

Realizátor sběru dat

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

Termín sběru dat

7. července až 10. listopadu 2015

Návratnost

41,7 %

Způsob sběru dat

Standardizovaný rozhovor face-to-face s využitím počítače (CAPI), vyplňování standardizovaného dotazníku na papíře (PAPI), vyplňování standardizovaného dotazníku na internetu (CAWI)

Jednotka analýzy

Domácnost a jednotlivec

Obsah a struktura datového souboru

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 a proměnných vztahujících se k průběhu dotazování

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

1. České panelové šetření domácností

České panelové šetření domácností (CHPS) je výběrové šetření, v němž je 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 listopadu 2015, bylo dotázáno 5 159 domácností. Do konce roku 2018 budou realizovány další 3 vlny sběru dat s roční periodicitou.

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, charakterizovat proces 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">• Vzdělanostní 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">• Vzdělanostní 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 organizuje 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 věd 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 spolupracuje na kontrole a čištění dat, zpracovává konečné datové soubory a zajišťuje jejich uložení v Českém sociálně vědním datovém archivu Sociologického ústavu AV ČR, v.v.i.

Sběr dat realizují 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, spolupracují na kontrole a čištění dat a poskytují 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. Do výzkumu jsou zařazeny domácnosti, jejichž členové jsou schopni zúčastnit se dotazování, tj. alespoň někteří z nich hovoří česky nebo slovensky a v účasti k dotazování jim nebrání 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, které 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 vymezujícím 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 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.

3. Sběr dat

3.1. Školení tazatelů

Pro práci na výzkumu byli vybíráni především tazatelé s dlouholetou praxí práce pro danou agenturu, se zkušeností s výzkumem na náhodně vybraném vzorku a CAPI dotazováním. Všichni tazatelé pracující na výzkumu prošli speciálním školením. Hromadná jednodenní školení probíhala ve vybraných městech České republiky v průběhu června 2015. Několik tazatelů bylo doškoleny individuálně.

Školení tazatelů se 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 (souběžné vyplňování CAPI a PAPI instrumentů),
- vybrané specifické momenty dotazování (např. zjišťování vztahů mezi členy domácnosti),
- 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.

Školení tazatelů předcházelo „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 období od 24. dubna do 19. května 2015 se 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 pilotním šetření bylo dotázáno 302 domácností vybraných postupem obdobným k proceduře využívané pro hlavní sběr.

3.3. Avizování výzkumu

Před každou etapou sběru dat byly 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.

3.4. Průběh terénních prací

Terénní práce na sběru dat pro první ročník výzkumu probíhaly v období od 7. července do 10. listopadu 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 se u hůře dosažitelných domácností maximalizovala šance kontaktu rozložením návštěv tazatele napříč delším obdobím. Kvůli nižší úspěšnosti kontaktu během letních prázdnin (červenec, srpen), byly cca 2 třetiny domácností dotázány od září do listopadu 2015.

Z vydaných 13 356 adres bylo dotázáno a kontrolou úspěšně ověřeno 5 159 domácností (41,7 % z adres splňujících požadavky na dotázání, podrobnosti v *tabulce 3*).

Na sběru dat se alespoň jednou dotázanou domácností podílelo 308 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 zbývající tazatele. V průměru na jednoho tazatele připadá cca 17 dotázaných domácností.

Tabulka 3: výsledky kontaktů ve vzorku domácností

vydané adresy	13 356 ^{*)}	
splňuje požadavky na dotázání („eligible“)	12 382	
dotázáno, ověřeno pomocí kontrol	5 159	41,7 %
odmítnutí	4 517	
dotazování nedokončeno nebo domluveno na termín po skončení terénu	162	
nezastižení (celé domácnosti nebo dospělého člena)	2 392	
dotázáno, vyřazeno na základě kontrol	152	
nesplňuje požadavky na dotázání („ineligible“)	974	7,3 %
neschopnost se zúčastnit - žádný člen nemluví česky	11	
neschopnost se zúčastnit - zdravotní stav, stáří	260	
nenalezená adresa	300	
adresa není obydlený byt	403	

^{*)} 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.

3.5. Kontaktování domácností

Při pokusu o kontakt s domácností měli tazatelé za úkol 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. Po prvním kontaktu mohl tazatel s domácností komunikovat telefonicky. Jestliže se tazateli nepodařilo domácnost zastihnout, mohl zanechat ve schránce navštívenku s datem a časem návštěvy a svým kontaktem, aby domácnost měla případně možnost se s tazatelem sama spojit.

Oslovené domácnosti mohly kontaktovat přímo realizátory a výzkumný tým na telefonních číslech a e-mailových adresách v avizním dopise a informačním letáku. Na webu výzkumu www.promeny-ceskespolecnosti.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 zobrazuje *tabulka 4*.

Od září 2015 bylo přijato několik dalších opatření na zvyšování šance kontaktu s dosud nezastiženými domácnostmi a konverze 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. Dotazníky na webu vyplnilo 589 domácností (11 % z celkového počtu dotázaných domácností). Úspěšnost CAWI konverze činila 9 %.

Tabulka 4: důvody odmítnutí účasti ve výzkumu

nezájem / nedůvěra k výzkumům obecně	2161	47,8 %
nedostatek času (členové domácnosti stále zaneprázdnění)	510	11,3 %
jiné (např. domácnost sama uvědomila agenturu, že se nechce zúčastnit)	455	10,1 %
nezájem o téma tohoto výzkumu	268	5,9 %
obava z nedostatku anonymity / zneužití informací / poskytování osobních údajů	266	5,9 %
některý člen domácnosti „zakázal“ účast	222	4,9 %
bez zdůvodnění	218	4,8 %
nedůvěra k tomuto výzkumu (nemá dost informací, vadí mu téma apod.)	138	3,1 %
špatné načasování, výjimečná situace v domácnosti (právě odjíždějí, nemoc, návštěva atd.)	118	2,6 %
předchozí špatná zkušenost s výzkumy	74	1,6 %
strach z tazatele (strach otevřít, strach z podomního prodeje)	49	1,1 %
nedostatečná motivace – malá odměna apod.	25	0,6 %
v nedávné minulosti na nějaký výzkum odpovídali	13	0,3 %

3.6. Monitorování terénních prací

Tazatelé zaznamenávali dny a výsledky návštěv na jednotlivých adresách, které 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 posledních dvou měsících se intenzita telefonických kontaktů zvýšila na minimálně dva za týden.

3.7. Rozhovor v domácnosti

Tazatelé usilovali o zajištění přítomnosti co nejvíce členů domácnosti během své návštěvy. Do domácností se mohli tazatelé opakovaně vracet, aby dokončili rozhovory, dotázali další členy domácnosti nebo vyzvedli vyplněné PAPI dotazníky.

Průměrná doba strávená tazatelem v domácnosti činila 110 minut. V úvodní části návštěvy představili tazatelé výzkum, způsob výběru domácnosti, požadavky na respondenty či systém odměňování za účast. Vysvětlovali rovněž způsob vyplnění PAPI dotazníků a časových snímků, aby bylo možné dosáhnout efektivního souběhu vyplňování instrumentů, tedy vyplňování PAPI dotazníků dalšími členy domácnosti, zatímco probíhalo CAPI dotazování.

Ke každému PAPI dotazníku předal tazatel respondentům propisovací tužku. Pro zajištění co nejhladšího průběhu dotazování rozdávali tazatelé mladším dětem v domácnosti pastelky a omalovánky, kterými se mohli zabavit během rozhovoru tazatele 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 PAPI dotazníků a časových snímků a případně vyzval respondenty k opravě či doplnění. Zúčastněným domácnostem mohli předat tazatelé kartičku s poděkováním.

3.8. Odměny pro respondenty a tazatele

Domácnosti získaly za účast ve výzkumu finanční odměnu odstupňovanou podle počtu členů domácnosti (od 500 Kč do 1000 Kč) a zohledňující počet vyplněných instrumentů (100 Kč navíc k základní odměně, pokud byly vyplněny všechny možné dotazové instrumenty). Respondentům byla nabídnuta možnost převést svoji odměnu na charitativní účely. Před začátkem sběru dat druhého ročníku výzkumu (červen 2016) bylo navíc náhodně vybráno 16 domácností zúčastněných v prvním roce, které získaly poukazy na nákup zboží.

Tazatelé byli odměňováni podle počtu vyplněných instrumentů v domácnostech a získávali bonusové složky odměny za včasné splnění tazatelského úkolu v přiděleném výběrovém bodě.

3.9. Kontrola práce tazatele

K ověření realizace rozhovoru sloužily dva mechanismy:

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

Alespoň jeden typ kontroly se uskutečnil ve všech domácnostech dotázaných tazatelem (kontrola náslechů u 82 % dotázaných domácností, telefonická/e-mailová/korespondenční kontrola u 94 % domácností). Pokud v rámci jedné kontroly vznikly pochybnosti o správné realizaci dotazování, byl proveden další typ kontroly. Druhou kontrolou prošly také ostatní domácnosti tazatele, u nějž byl zaznamenáno pochybení. Při podezření na provedení rozhovoru s jinou než předepsanou domácností provedl osobní kontrolu v terénu pracovník tazatelské sítě.

Mimo uvedené dva druhy kontroly realizace rozhovoru probíhala 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 byly uskutečněny na základě unikátních přístupových kódů doručených poštou přímo do vybraných domácností, a to bez zprostředkování tazatelem.

3.10. Udržování kontaktu s domácnostmi

V prosinci 2015 bylo domácnostem zúčastněným v prvním ročníku zasláno elektronicky přání k Vánocům společně s informací o úspěšném zakončení sběru dat.

4. Dotazové instrumenty

4.1. Typy instrumentů

Data byla sbírána pomocí standardizovaných dotazníků formou osobního rozhovoru s využitím počítače (CAPI) nebo samo-vyplňování (PAPI). Pro účely dotazování po internetu byly všechny instrumenty převedeny také do webové formy (CAWI).

CAPI dotazník pro domácnost vyplňoval 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. Průměrná délka CAPI dotazníku pro domácnost byla 16,5 minuty.

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 je 6,5 %. Průměrná délka CAPI individuálního dotazníku činila 28 minut, CAPI proxy dotazníku 8 minut.

Pro CAPI domácnostní, individuální a proxy dotazování byly připraveny **karty** k vybraným otázkám.

PAPI dotazníky pro dospělé byly předávány všem členům domácnosti starším 18 let, **PAPI dotazníky pro děti** byly určeny dětem od 10 do 17 let². Ačkoli dotazníky pro děti a dospělé obsahují některé společné otázky, jejich obsah se do značné míry liší. Dětské PAPI dotazníky byly 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. Dětské PAPI dotazníky měly být vyplňovány bez dohledu rodičů.

Časový snímek předchozího dne měl dvě varianty s různými kategoriemi využívání času, a to **pro děti** od 10 a 17 let a **pro dospělé**. Rodiče mohli mladším dětem pomoci s časovým snímkem, pokud měly děti s vyplňováním potíže.

² Agentury realizující sběr dat upravily formát symbolů pro kódování odpovědí v dotaznících s ohledem na svůj postup při pořizování dat (MEDIAN: scanování, STEM/MARK: ruční pořizování).

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. V typickém případě vyplňovali respondenti papírové instrumenty během CAPI rozhovoru tazatele s jiným členem domácnosti. V některých případech tazatel předal PAPI 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 PAPI 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 vracet.

Protože nebyla stanovena žádná hierarchie vyplňování dotazových instrumentů, může u každého respondenta existovat jakákoli kombinace vyplněných dotazníků, které byly určeny osobám v jeho věku (s výjimkou kombinace CAPI individuálního dotazníku a CAPI proxy dotazníku).

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 je zařazeno také 212 částečně dotázaných domácností, v nichž byl vyplněn alespoň CAPI dotazník pro domácnost.

Tabulka 5: počty vyplněných instrumentů

CAPI pro domácnost, proměnné na úrovni domácnosti	5 159
CAPI pro domácnost, individuální proměnné	13 049
<i>z toho:</i>	
dospělí	10 469
děti 10 až 17 let	1 077
děti do 10 let	1 503
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>	
10-14 let	306
15-17 let	560
časový snímek dospělí	7 955
časový snímek děti	804

Tabulka 6: míra vyplnění instrumentů celkem a podle typu domácnosti (hhtype)

	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	68,0%	99,6%	71,9%	63,1%	85,3%	52,1%	67,9%	50,0%
CAPI individuální včetně proxy	72,6%	99,6%	75,7%	69,5%	88,6%	58,5%	74,0%	54,8%
PAPI dospělí	77,7%	77,7%	77,7%	75,3%	83,5%	75,3%	81,2%	73,3%
PAPI děti	80,4%	-	-	-	80,3%	-	79,2%	90,0%
časový snímek dospělí	76,0%	76,9%	76,2%	72,2%	81,3%	74,7%	79,1%	70,6%
časový snímek děti	74,7%	-	-	-	75,1%	-	73,5%	82,7%

Datový soubor obsahuje proměnné na úrovni domácnosti s počtem a podílem jednotlivých vyplněných dotazníků (viz oddíl 9. *Proměnné vztahující se k průběhu a podmínkám dotazování*).

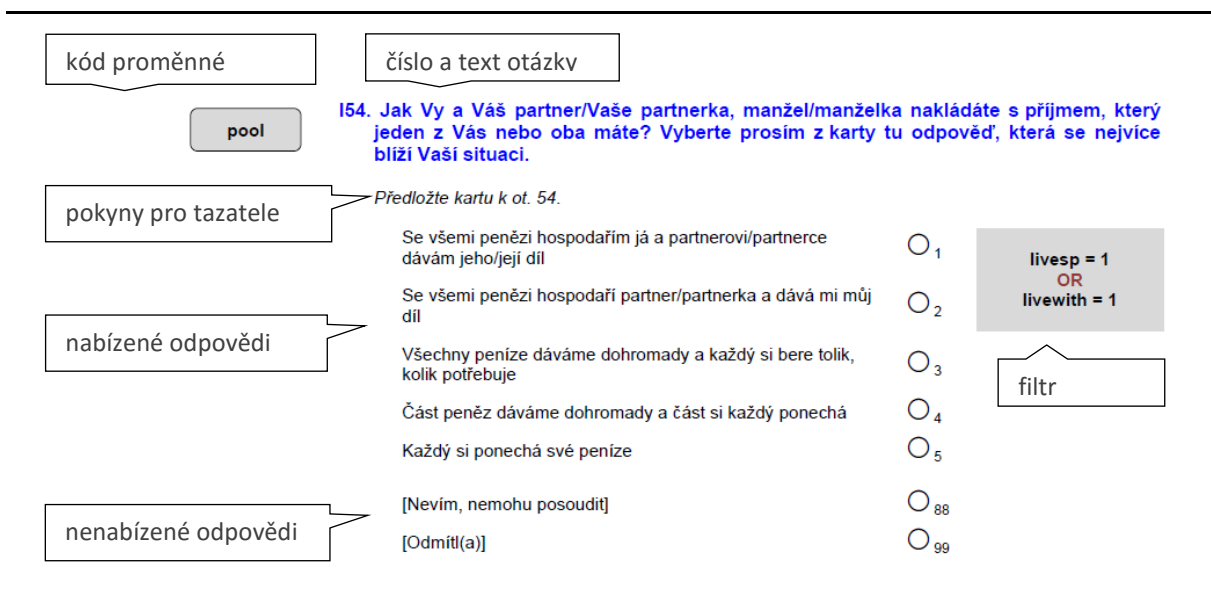
4.4. Dokumentace instrumentů

Labels proměnných v datovém souboru zkráceně vystihují text otázky. Plné znění otázek včetně variant odpovědí je dostupné v jednotlivých dotazových instrumentech.

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 [...].

Schéma 1: popis náležitostí publikovaných CAPI dotazníků



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).

Schéma 2: 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"/>

5. Obsah a struktura datového souboru

Datový soubor z prvního ročníku CHPS obsahuje proměnné z CAPI a PAPI dotazových instrumentů. Proměnné z časových snímků pro dospělé a děti nejsou v datovém souboru zahrnuty. 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 7*. Proměnné, které byly dotazovány ve více instrumentech, jsou v datovém souboru zařazeny podle umístění ve dříve řazeném instrumentu. Výjimkou jsou proměnné *health* a *hamper*, dotazované v CAPI proxy dotazníku a v PAPI dotazníku pro dospělé, které jsou umístěny podle pořadí v 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).

Labels 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 7: 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
PAPI dotazník pro děti 10-14 let	C + číslo otázky v labelu proměnné
stratifikační postavení	
místo bydliště domácnosti	
charakteristiky dotazování	
časy dotazování	
vyplnění dotazníků v domácnosti	
indikace vyplnění instrumentů a váhy	

5.2. Struktura souboru

Datový soubor obsahuje 13 049 případů, které reprezentují členy dotázaných domácností. 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.

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). Schéma struktury datového souboru ilustruje *tabulka 8*.

Tabulka 8: schéma struktury 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 datovém souboru 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), údaje sloužící ke kontrole v průběhu dotazování (např. *namech*) a proměnné vztahující se k 11. a 12. členům domácnosti (vyšetřené domácnosti měly maximální 10 členů).

Proměnné nezařazené do datového souboru shrnuje *tabulka 9*.

Tabulka 9: proměnné nezařazené do datového souboru

name	křestní jméno (případně přezdívka)	CAPI dotazník pro domácnost
rel11-rel12	vztahy k členům domácnosti pno 11 a 12	CAPI dotazník pro domácnost
pidch11-pidch12	identifikační číslo dítěte pno 11 a 12	CAPI dotazník pro domácnost
namech	kontrola zadání správného jména	CAPI individuální dotazník
sexch	kontrola zadání správného pohlaví	CAPI individuální dotazník
birthmch	kontrola zadání správného měsíce narození	CAPI individuální dotazník
birthych	kontrola zadání správného roku narození	CAPI individuální dotazník
educch	kontrola zadání správného vzdělání	CAPI individuální dotazník
stdych	kontrola zadání správného studijního stavu	CAPI individuální dotazník
estatch	kontrola zadání správného ekonomického postavení	CAPI individuální dotazník
mstatch	kontrola zadání správného rodinného stavu	CAPI individuální dotazník
sname(x)	název školy studované v současnosti	CAPI individuální a proxy dotazník
sstr(x)	ulice školy studované v současnosti	CAPI individuální a proxy dotazník
smun(x)	obec školy studované v současnosti	CAPI individuální a proxy dotazník
sprog(x)	obor studovaný v současnosti	CAPI individuální a proxy dotazník
wrkps	popis pracovní činnosti respondenta	CAPI individuální a proxy dotazník
eduas11-eduas12	aspirované vzdělání dítěte pno 11 a 12	CAPI individuální dotazník
eduex11-eduex12	očekávané vzdělání dítěte pno 11 a 12	CAPI individuální dotazník
liwhy	jiný důvod, proč respondent nežil do 16 let převážně s oběma biologickými rodiči	CAPI individuální dotazník
mwrkps	popis pracovní činnosti matky respondenta v době, kdy respondentovi bylo 16 let	CAPI individuální a proxy dotazník
fwrkps	popis pracovní činnosti otce respondenta v době, kdy respondentovi bylo 16 let	CAPI individuální a proxy dotazník
mstreet	ulice předchozího bydliště	CAPI individuální dotazník
mcounty	okres předchozího bydliště	CAPI individuální dotazník
mmunic	obec předchozího bydliště	CAPI individuální dotazník
mpsc	PSČ předchozího bydliště	CAPI individuální dotazník
papiname	křestní jméno vyplněné v PAPI dotazníku	PAPI dospělí a děti
papiage	věk vyplněný v PAPI dotazníku	PAPI dospělí a děti
papisex	pohlaví vyplněné v PAPI dotazníku	PAPI dospělí a děti
jobas	popis práce, kterou chce dítě v budoucnu vykonávat	PAPI děti
jobex	popis práce, kterou si dítě myslí, že bude v budoucnu vykonávat	PAPI děti

6. Identifikační proměnné

Domácnosti jsou odlišeny unikátními identifikačními čísly (*hid*), která jsou společná všem členům jedné domácnosti. Unikátní identifikační číslo osoby (*pid*) vzniká jako kombinace identifikace domácnosti (*hid*) a pořadí osoby v rámci dané domácnosti (*pno*). Unikátní identifikační číslo osoby (*pid*) určuje respondenta individuálního a PAPI dotazníků a identifikuje osobu, za niž byl vyplněn CAPI proxy dotazník. Identifikační číslo respondenta CAPI proxy dotazníku obsahuje proměnná *pidpro*, identifikační číslo respondenta CAPI dotazníku pro domácnost obsahuje proměnná *pidhr*.

K jednotlivým osobám jsou přiřazena identifikační čísla dalších členů domácnosti, kteří mají k osobě vybraný vztah (*tabulka 11*).

Tabulka 10: identifikační proměnné jednotlivce/domácnosti

hid	unikátní identifikační číslo domácnosti
pid	unikátní identifikační číslo osoby
pno	pořadí osoby v domácnosti
pidhr	identifikační číslo osoby, která vyplnila dotazník pro domácnost
pidpro	identifikační číslo osoby, která vyplnila proxy dotazník

Tabulka 11: 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 vážících se k druhému pohlaví, než je pohlaví biologického rodiče.

7. 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. Individuální generované proměnné odvozené z proměnných v CAPI individuálním nebo proxy dotazníku mají hodnoty pro respondenty, kteří vyplnili daný instrument.

Tabulka 12: generované proměnné na úrovni 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 ($age \geq 18$)

Na úrovni domácnosti bylo zkonstruováno několik ukazatelů struktury domácností, a to proměnná *hhtype* a sada proměnných *hhtype2* až *hhtype6* (tabulky 13 a 14). 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 13: 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 14: 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 15: 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>)
agecat	věk kategorizovaný po 10 letech
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
livemo^{*) **)}	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>)
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>)
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
mrenum	počet zvolených důvodů k přestěhování (<i>mwork, mfam, mhous, melse</i>) respondentů CAPI individuálního dotazníku

^{*)} 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/kteřá žije mimo domácnost).

^{**)} 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.

8. Stratifikační postavení

Hodnotám proměnných *isco08*, *fisco08*, *misco08*, *aisco08* a *eisco08* byly přiřazeny hodnoty ukazatele ISEI (International Socio-Economic Index)³.

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)⁴.

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 16: 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)

³ 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>.

⁴ Harrison, Eric. „European Socio-economic Classification.“ <http://ekharrison.weebly.com/european-socio-economic-classification-esec.html> <February 5 2015>.

9. Proměnné vztahující se k průběhu a podmínkám dotazování

Datový soubor obsahuje 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*).

Tabulka 17: místo bydliště domácnosti

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

Data obsahují indikátory umožňující rozlišit mód dotazování (CAPI = osobní rozhovor s tazatelem a vyplnění papírových dotazníků, CAWI = vyplnění všech instrumentů přes internet) a agenturu, která dotázala danou domácnost. Doplněno je několik indikací typu vyplněných instrumentů, tazatelem zaznamenaný důvod provedení proxy rozhovoru a odvozený vztah mezi proxy respondentem a osobou, za niž vypovídá.

Tabulka 18: charakteristiky dotazování

qmode	zda se domácnost zúčastnila osobního rozhovoru nebo prostřednictvím CAWI
agen	agentura, která provedla dotazování domácnosti
iorder	pořadí vyplnění CAPI individuální dotazníku v domácnosti (pouze pro respondenty, s nimiž byl přímo vyplněn CAPI individuální dotazník)
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 19: 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 individuální dotazníku
iendm	měsíc ukončení CAPI individuálního dotazníku
iendd	den ukončení CAPI individuálního dotazníku
pstartm	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 počet osob v daném věku v domácnosti. Pokud v domácnosti nebyla žádná osoba z věkové skupiny, pro niž byl daný instrument určen, hodnota je prázdná (system missing).

Tabulka 20: 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

10. 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 souhlasili s využitím údajů jen pro uvedené účely.

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 nemají vazbu na výběrové body.

11. Kontrola, zpracování a čištění dat

11.1. Kontroly dat

Do CAPI instrumentů bylo zakomponováno několik mechanismů na podporu zajištění správnosti a konzistence dat. V CAPI dotazníku pro domácnosti 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.

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 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 definice 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 nevznikly posléze v procesu manipulace s daty. Proměnné s odpovědí ve formě roku určité události (např. *madres*) prošly kontrolou konzistence výpovědi. Ačkoli byly během procesu kontroly identifikovány vzájemně nekonzistentními odpovědi 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.

Během monitorování terénních prací byla také kontrolována shoda distribucí mezi agenturami zajišťujícími sběr dat. V případech rozdílů byly zjišťovány možné příčiny a zajišťována harmonizace v postupu tazatelů.

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.

11.2. 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 proměnné *age* upravena podle hodnoty proměnné *papiage*.

Ze zpracování byly vyřazeny PAPI dotazníky, v nichž chybělo více než 25 % odpovědí („bez odpovědi“ nebo „nejednoznačná odpověď“), a časové snímky, v nichž chyběly odpovědi ve více než 6 časových úsecích. Vyřazeny byly rovněž papírové instrumenty, které kvůli chybějícím identifikačním kódům nemohly být propojeny s ostatními dotazníky.

CAPI dotazníky kvůli většímu podílu neplatných odpovědí („nevím“, „odmítl(a)“) vyřazovány nebyly.

11.3. Tvorba datového souboru

Realizátoři předávali výzkumnému týmu oddělené datové soubory s proměnnými z jednotlivých instrumentů (CAPI dotazník pro domácnost byl 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*).

11.4. 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* a *birthy* byly začištěny podle respondenty opravených hodnot z druhého ročníku výzkumu (CAPI dotazník pro domácnost ve druhé vlně nabízel respondentům možnost opravy údajů zapsaných v prvním roce).

Proměnné *mstat*, *estat*, *educ*, *std*, *stdch* a *rel1* až *rel12o2* byly individuálně začištěny podle hodnot návazných proměnných (např. respondent uvedl v otevřené odpovědi proměnné *stypc*, že ve skutečnosti nestuduje), hodnot zadaných ve druhé vlně a hodnot návazných proměnných ve druhé vlně s přihlédnutím k celkové logice jednotlivých odpovědí.

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).

11.5. Kódování otevřených a polootevřených otázek

V polootevřených a otevřených otázkách *rel, lang1, lang2, dwelob, edusec, eduprim, liwhy, rework, refam, rehous, reelse, mdwelob, cborn, rgborn, citoc, 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 (*wrkps, fwrkps, mwrkps, jobas, jobex*) jsou zakódovány podle klasifikace CZ-ISCO.

11.6. Imputace hodnot

V CAPI a 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. Imputace je v datech z časových snímků značena.

11.7. Úpravy počtu případů podle vlny 2

Po ukončení sběru dat ve druhém roce výzkumu bylo do dat z první vlny doplněno 10 osob, které žily v dotázaných domácnostech již v prvním roce, ale nebyly tazatelem zaznamenány. Zároveň bylo z dat vyřazeno 41 domácností, u nichž opakování výzkumu odhalilo závažná pochybení v realizaci rozhovorů v prvním roce (započítáno v *tabulce 3: výsledky kontaktů ve vzorku domácností*).

11.8. Neplatné odpovědi/hodnoty

Všechny otázky v CAPI dotaznících dovolovaly 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“. V otázkách na politické znalosti (*knowkal* až *knowdsss*) 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 popsanych v *tabulce 21*.

Kódy neplatných hodnot jsou v SPSS datových souborech nastaveny jako user-missing. Jako user-missing jsou dále nastaveny kódy 96 / 999 996 v otázkách *relparnl*, *fowny*, *fcoopy*, *symods* až *symano*. Ačkoli se jedná o respondentům nabízené platné možnosti, značí odpověď, která se nachází mimo hlavní škálu variant odpovědí.

Tabulka 21: 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>
99 / 999 999	CAPI	odmítl(a) – nebylo nabízeno <i>V CAWI mutaci dotazníku uvedeno mezi možnostmi ve znění „Nechci uvést“.</i>
	PAPI	bez odpovědi
98 / 999 998	CAPI	nebylo dotázáno v důsledku technické chyby při dotazování nebo navolení chybných filtračních hodnot
	CAPI	<i>part</i> , <i>sumchild</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>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)

12. Specifika vzorku

V datech CHPS jsou ve srovnání s údaji ze Sčítání lidu, domů a bytů 2011⁵ 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.

Dílčí skupiny respondentů vymezené vyplněnými instrumenty se z hlediska složení vzájemně liší, přítomné jsou zejména odchylky mezi vzorkem všech členů domácnosti a vzorky respondentů, kteří vyplnili jednotlivé individuální dotazníky. 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 socio-demografických segmentů populace (ženy, starší lidé atp.) ve srovnání se souborem všech členů domácnosti.

⁵ Sčítání lidu, domů a bytů 2011, Český statistický úřad. Výsledky na: <https://www.czso.cz/csu/sldb/domacnosti>

13. Post-stratifikační váhy

13.1. 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. 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 výběr těchto jednotlivých skupin respondentů slouží binární proměnné *hh_hh* až *child_papi_diary*.

Binární proměnné *hh_hh* až *child_papi_diary* 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 22*, konstrukci dílčích vzorků respondentů ilustruje *tabulka 25*.

Tabulka 22: indikátory skupin respondentů a příslušné váhy

hh_hh	dotázaná domácnost (člen domácnosti, který vyplnil CAPI dotazník pro domácnost)
W_hh_hh	váha pro domácnosti
hh_indi	člen dotázané domácnosti
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
diary_child	člen domácnosti, který vyplnil časový snímek pro děti
W_diary_child	váha pro respondenty časového snímku pro děti
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
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
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
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

13.2. 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 předem stanovena na hodnotu 4. Fixace však nemusela být využita, protože žádná z vážících proměnných tuto hodnotu nepřesáhla. R script pro provedení vážení je dostupný jako samostatný soubor.

13.3. 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 jsou data z Výběrového šetření pracovních sil (VŠPS) 2015, Sčítání lidu, domů a bytů (SLDB) 2011 a údaje Českého statistického úřadu o věkovém složení obyvatel k 1. 7. 2015.

Tabulka 23: 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. 2015 (ČSÚ) pohlaví, věk, vzdělání: VŠPS 2015
3	pohlaví + kraj bydliště (NUTS 3)	kraj, pohlaví: Věkové složení mužů k 1. 7. 2015 a Věkové složení žen k 1. 7. 2015 (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 24*). Kontrolní data využitá pro odvození teoretických distribucí jsou dostupná jako samostatný MS Excel soubor.

Tabulka 24: 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	1	$10 \geq age \leq 34$	věk3	1	$15 \geq age \leq 34$
	2	$18 \geq age \leq 34$	(ageR2)	2	$35 \geq age \leq 54$	(ageR3)	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)						

13.4. 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 25: 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	pro- měnné vážení	kat. věku
hh_hh	x								5159	1	-
hh_indi	x								13049	2	1
indi	&	&							7118	2	1
indi_proxy	&	+	+						7605	2	1
papi_adult	&			&					8131	2	1
papi_child	&				+	+			866	3	-
adult_child	&			+	+	+			8997	2	2
adult_child15	&			+	+				8437	2	3
indi_child15	&	+			+				7424	2	3
diary_adult	&						&		7955	2	1
diary_child	&							&	804	3	-
indi_papi	&	&		&					6339	2	1
indi_diary	&	&					&		6239	2	1
adult_papi_diary	&			&			&		7589	2	1
indi_papi_diary	&	&		&			&		5989	2	1
child_papi_diary	&				+	+		&	778	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

14. Přehled materiálů k prvnímu ročníku CHPS

Tabulka 26: přehled materiálů k prvnímu ročníku CHPS

Kontaktní materiály	
Avizní dopis pro respondenty	varianta MEDIAN a STEM/MARK
Informační leták pro respondenty	varianta MEDIAN a STEM/MARK
Dopis obcím	varianta MEDIAN a STEM/MARK
Dopis městům	
Dopis krajským ředitelstvím Policie ČR	
Navštívenka	varianta MEDIAN a STEM/MARK
Děkovka	varianta MEDIAN a STEM/MARK

Dotazové instrumenty	
CAPI dotazník pro domácnost	
CAPI individuální dotazník	
CAPI proxy dotazník	
PAPI dotazník pro dospělé *)	varianta MEDIAN a STEM/MARK
PAPI dotazník pro děti 15-17 let *)	varianta MEDIAN a STEM/MARK
PAPI dotazník pro děti 10-14 let *)	varianta MEDIAN a STEM/MARK
Časový snímek pro dospělé	
Časový snímek pro děti	
Karty ke CAPI dotazníku pro domácnost	
Karty ke CAPI individuálnímu dotazníku	
Karty ke CAPI proxy dotazníku	

Datový soubor	
Datový soubor CHPS W1 (2015)	
Seznam proměnných v datovém souboru	
SPSS syntax pro konstrukci ukazatelů ISEI a ESeC	v anglickém jazyce
R script pro provedení vážení	v anglickém jazyce
Kontrolní distribuce využití při vážení	v anglickém jazyce

*) Dotazník distribuovaný respondentům neobsahoval kódy jednotlivých proměnných.