

Data z pilotní covidové studie výzkumu „Současná česká rodina“ jsou volně dostupná pro výzkumné účely

České i mezinárodní odborné komunitě jsou od léta 2021 k dispozici unikátní data o soudobé české rodině. Jde o data z pilotní covidové studie výzkumu „Současná česká rodina“. Vzorek s více než 1 300 respondenty dotázanými od prosince 2020 do února 2021 byl nově zpřístupněn v datovém archivu projektu GGP na adrese www.ggp-i.org. Věříme, že tato data, jejichž aktuálnost je zdůrazněna probíhající pandemií nemoci COVID-19, najdou v české odborné komunitě široké užití u stávajících i nových uživatelů.

Kontext a historie výzkumu

Český projekt „Současná česká rodina“ (<https://ggp-cz.fss.muni.cz>) je součástí mezinárodní výzkumné infrastruktury Generations and Gender Programme (GGP), která poskytuje komparativní longitudinální data o populačním chování, zejména o rodině a životní dráze, a to s cílem přispět k řešení aktuálních společenských výzev i výzev veřejných politik. Jednou z hlavních součástí výzkumného programu GGP je panelové výběrové šetření „Generations and Gender Survey“ (GGS). Projekt „Současná česká rodina“ je českou částí právě tohoto výběrového šetření.

Výzkum je zaměřen na dospělé populaci zúčastněných zemí (obvykle, ale ne vždy ve věku 18 až 79 let). GGS používá „cyklický design“, který lze charakterizovat jako panelový výzkum opakovaný v pravidelných cyklech vždy na novém vzorku respondentů. Interval mezi panelovými rozhovory je tři roky a každý cyklus může zahrnovat tři vlny dotazování. Dotazník GGS zahrnuje širokou škálu témat, zaměřuje se mj. na osobní historie plodnosti a partnerství, vztahy, rozdělení domácích prací, rovnováhu mezi prací a rodinou, přechod do dospělosti, mezigenerační výměnu, ekonomické aktivity, zdraví a plodnost.

První kolo GGS bylo zahájeno v polovině první dekády 21. století a zúčastnilo se ho dvacet zemí; mezi nimi byla i Česká republika, kde první dotazování proběhlo v roce 2005. Druhé kolo, v České republice zaštitěné projektem „Současná česká rodina“, bylo zahájeno na konci roku 2019, přičemž se ho účastní více než třicet zemí.

Koordinátorem GGP je v současnosti profesorka Anne H. Gauthier z Nizozemského mezioborového demografického institutu (Netherlands Interdisciplinary Demographic Institute, NIDI). Na řešení české části projektu se podílí Masarykova univerzita (Fakulta sociálních studií), Univerzita Karlova (Přírodovědecká fakulta) a Výzkumný ústav práce a sociálních věcí, v. v. i.; realizátorem sběru dat je agentura SC&C. Vedoucím projektu je prof. Martin Kreidl z Masarykovy univerzity. Projekt je podporován Technologickou agenturou ČR (projekt č. TL03000338, program ÉTA).

Okolnosti vzniku pilotní covidové studie výzkumu Současná česká rodina

Pilotní covidová studie výzkumu „Současná česká rodina“ se uskutečnila před započítím sběru dat pro hlavní dotazníkové šetření, a to jako jeho pilotáž, ale i jako reakce na nepříznivý vývoj pandemie COVID-19 na podzim roku 2020 a podmínky financování projektu.

Původní harmonogram výběrového šetření počítal s tím, že sběr dat bude proveden mezi 6 000 respondenty ve věku 18 až 69 let v období od října 2020 do listopadu 2021. Podle mezinárodně sdílené metodiky projektu GGP využívá hlavní GGS výzkum k výběru respondentů víceetapovou metodu pravděpodobnostního výběru. Vzhledem k tomu, že v České republice není pro výzkumné účely k dispozici žádný registr obyvatel, byl za oporu výběru zvolen „Registr územní identifikace, adres a nemovitostí“ (RÚIAN, spravovaný ČSÚ). RÚIAN nicméně obsahuje pouze adresy (případně počet bytových jednotek na

dané adrese), nikoliv jména rezidentů. Povaha tohoto registru tak vyžaduje, aby každou vybranou adresu/domácnost kontaktoval tazatel, sestavil seznam všech způsobilých členů vybrané domácnosti a využil tzv. metodu posledních narozenin k výběru respondenta z této domácnosti pro rozhovor.

Jakkoli je u výzkumů podobného charakteru obvyklé provádět pilotáž, pilotní studie výzkumu „Současná česká rodina“ je s více než 1 300 respondenty neobvykle rozsáhlá. Její rozsah, design i načasování byly důsledkem dvou okolností.

První okolností byl neočekávaný vývoj pandemie COVID-19 v posledních měsících roku 2020. Zatímco v letních měsících se zdálo, že je pandemie v zásadě pod kontrolou, v září 2020 se situace začala zhoršovat, přičemž negativní vývoj se výrazně urychlil v průběhu října. Česká vláda reagovala 22. října 2020 zavedením druhé vlny přísných celostátních mimořádných a ochranných opatření, která výrazně omezila sociální mobilitu, osobní kontakty a která zakázala i mnoho druhů ekonomických aktivit. Ačkoliv některé obchodní aktivity zůstaly povoleny, výzkumné rozhovory tazatelů byly z právního hlediska hraničním případem. Dotazování v terénu se tak ukázalo nejen jako obtížně proveditelné, nýbrž vzhledem k panujícím obavám i potřebám omezení osobních kontaktů také jako nevhodné a rizikové. Původní plán založený na využití tazatelů k rekrutaci respondentů jsme proto přepracovali a pokusili se tento způsob náboru nahradit nábohem poštou. Tento postup se však ukázal jako velmi neefektivní. Uvedenou metodu náboru jsme zpočátku testovali na dílčím vzorku 3 012 adres. Z výběrového souboru jsme vybrali 21 084 adres. Tyto adresy byly rozděleny do čtyř dávek, které měly být postupně rozděleny tazatelům na základě vývoje míry návratnosti. Největší dávka obsahovala šestnáct adres v každé primární výběrové jednotce, zbývající dávky obsahovaly čtyři další adresy v každé

primární výběrové jednotce. Adresy z jedné z těchto menších dávek jsme použili k testování postupu náboru poštou. Z použitých adres se nám nicméně v období od 19. listopadu (kdy byly rozeslány první zvací dopisy) do 7. prosince podařilo získat pouze čtyřicet dokončených rozhovorů. Bylo proto evidentní, že nábor poštou nemůže zachránit hlavní terénní práci, zvláště když přítomnost tazatelů v terénu byla nutná i pro realizaci postupů náhodného výběru.

Druhou okolností stojící v pozadí vzniku covidového pilotu byly podmínky financování projektu „Současná česká rodina“. Finanční podmínky vyžadovaly, aby už v roce 2020 bylo získáno určité množství dat, na nichž by bylo možné následně založit výzkumnou zprávu, a aby tak již v tomto roce byla vynaložena určitá část rozpočtu projektu určená na vlastní sběr dat. Hlavní zdroj financování projektu zajišťuje Technologická agentura České republiky v rámci svého programu „ÉTA“, který vyžadoval spolufinancování ze strany přijímajících organizací. Toto spolufinancování poskytly všechny čtyři zúčastněné instituce, většinou formou částečných příspěvků na osobní náklady, nicméně značná část (1 mil. Kč) byla výslovně určena na pokrytí nákladů na terénní práce. Rozpočtová pravidla daná podmínkami VÚPSV (resp. Ministerstva práce a sociálních věcí jako zřizovatele) coby jednoho z přijímajících subjektů zde přitom určovala, že tato částka musela být vynaložena již v roce 2020. Jak již bylo uvedeno, z důvodů spojených s pandemií COVID-19 probíhaly terénní práce v hlavním GGS výzkumu velmi pomalým tempem. Hrozilo tak, že peníze na terénní práce vyhrazené na rok 2020 nebudou vyčerpány, což představovalo významné riziko ve vztahu s financující agenturou a spolufinancující institucí.

Tváří v tvář těmto dvěma podstatným okolnostem jsme proto (se souhlasem všech zúčastněných institucí) navrhli provedení obsáhlé pilotní GGS studie.

O datech covidové pilotní studie

Sběr dat pro pilotní GGS studii proběhl ve dvou vlnách, v prosinci 2020 a v dubnu 2021. Odborné komunitě jsou v tuto chvíli k dispozici pouze data z první vlny z prosince 2020; ke zpřístupnění dubnových dat by mělo dojít v období jara 2022.

Pro účely *první vlny v prosinci 2020* byl použit stejný dotazník, který byl připraven pro hlavní výběrové šetření GGS. Základem dotazníku byl přeložený dotazník sdílený v rámci mezinárodní komunity GGP. Toto jádro dotazníku obsahuje sady otázek rozdělených do devíti okruhů (demografie, životní dráha, fertilita, domácnost, rodinné zázemí a generační vazby, zdraví a duševní pohoda, práce, příjmy a postoje) a pro potřeby české větve projektu je doplněné o dalších čtyřicet národně specifických otázek zaměřených zejména na ekonomickou nejistotu a důsledky pandemie COVID-19 a na pracovní podmínky. Vyplnění dotazníku zabralo průměrně šedesát minut. Při sběru dat byla použita výhradně technika online dotazování pomocí počítače (CAWI).

Pilotní GGS studie není – na rozdíl od dat hlavního GGS výzkumu – založena na metodě náhodného výběru. Nepravděpodobnostní vzorek byl získán na základě kvótního výběru z tzv. „opt-in“ panelu SC&C tvořeného poloprofesionálními respondenty, kteří se účastnili studií provedených v letech 2012 až 2020 (v průměru se před účastí v této pilotní studii jednalo o jedenáct jiných výzkumů). Tito respondenti buď v nějakém předchozím šetření souhlasili s tím, že budou znovu osloveni pro další projekty, nebo se sami nahlásili agentuře jako dobrovolníci pro sběr dat. K účasti v pilotní GGS studii byla oslovena skupina více než 7 000 takových respondentů. Pro kontrolu složení vzorku byl použit jednoduchý kvótní výběr na základě pohlaví a věku (pět kategorií). Kvóty vycházely z údajů ČSÚ. Tato náborová strategie byla velmi účinná a během dvou týdnů dotaz-

ník vyplnilo více než 1 300 respondentů. Respondentům byla nabídnuta finanční odměna 500 Kč za vyplnění dotazníku.

Celkem bylo v první vlně získáno 1 327 respondentů tvořených osobami ve věku 18 až 69 let s pobytem v soukromých domácnostech bez ohledu na jejich národnost, státní příslušnost, jazyk nebo právní status. Sběr dat probíhal od 9. prosince 2020 do 9. února 2021. Do 14. prosince bylo získáno celkem 788 vyplněných dotazníků, do 21. prosince 1 153 a do konce roku 1 282. Poslední vyplněný dotazník byl získán 12. února 2021. Parametry získaných dat relativně dobře odpovídají struktuře obyvatelstva České republiky, vzhledem ke krátkému času vymezenému pro sběr dat se nicméně nepodařilo zcela naplnit použité kvóty (zejména u mladých mužů do třiceti let je v datech patrně nižší zastoupení).

V dubnu 2021 byla provedena *druhá vlna* navazující na prosincovou pilotní GGS studii. Vytvořen byl krátký dotazník složený částečně z replikovaných otázek z prosincové pilotní studie a částečně z nových otázek zaměřených zejména na vnímání a důsledky pandemie COVID-19. Vyplnění dotazníku zabralo průměrně patnáct minut. Při sběru dat byla i zde použita výhradně technika dotazování pomocí počítače (CAWI).

K účasti v této druhé vlně byli vyzváni všichni respondenti prosincové pilotní studie dne 4. dubna. Dotazování bylo dokončeno 19. dubna. Sběr dat tak probíhal převážně v době tzv. tvrdého lockdownu zavedeného podle uzavření všech škol také omezení pohybu v rámci okresu trvalého bydliště. Navazující druhé vlny se zúčastnilo celkem 1 201 respondentů, což vedlo k 9% míře úbytku. To je obecně pro panelové šetření v České republice i jinde velmi nízká míra úmrtnosti.

Dostupnost dat

Data z pilotní covidové studie výzkumu „Současná česká rodina“ jsou zdarma pří-

stupná veškeré vědecké veřejnosti v centrálním archivu GGP. Pro získání přístupu je nutné se zaregistrovat do „GGP User Space“ na adrese www.ggp-i.org. Následně je nutné vyplnit formulář žádosti o přístup k datům „Czech Republic – GGP 2020 – Czech Republic – Pilot“ a zaslat e-mailem podepsaný souhlas s podmínkami pro využívání dat.

Složka souborů ke stažení obsahuje kromě datových matic také soubor s dokumentací usnadňující orientaci v datech a upozorňující na některé provedené úpravy dat. Data jsou dostupná ve formátu SPSS. Jednotlivé datové matice pro každou

z vln obsahují vždy na počátku identifikační číslo (ID) daného respondenta. Toto číslo je unikátní pro každého respondenta a umožňuje spojit obě matice pro potřeby analýzy obou vln zároveň. Konkrétní logiku značení proměnných v jádrovém dotazníku i způsob vytvoření konstruovaných proměnných vysvětluje metodologický průvodce, který je k dispozici na adrese projektu GGP (<https://www.ggp-i.org/data/methodology/>).

*Tomáš Zvoníček, Martin Kreidl
Fakulta sociálních studií,
Masarykova univerzita, Brno*