

VYUŽITÍ ADMINISTRATIVNÍCH ZDROJŮ DAT PŘI VYMEZENÍ OBYVATELSTVA VE SČÍTÁNÍ LIDU 2021 V ČESKU

Robert Šanda¹⁾

THE USE OF ADMINISTRATIVE DATA SOURCES IN THE 2021 POPULATION CENSUS IN CZECHIA

Abstract

The article presents the administrative data sources used in the 2021 population census in Czechia and describes the main steps in data processing leading from raw input data to the final census population. Special attention is paid to signs-of-life analysis, which deals with the issue of overcoverage in the population register. The article ties in with a previous article (published in 2022) that discussed the use of registers in European population censuses and explained the process of combining data sources in the previous 2011 Czech population census.

Keywords: population and housing census, Czechia, administrative data sources, data quality, signs-of-life analysis
Demografie, 2023, **65(3): 120–135**

DOI: <https://doi.org/10.54694/dem.0323>

1. ÚVOD

Předkládaný článek navazuje na dříve publikovanou práci, která se věnovala problematice administrativních zdrojů dat ve sčítání lidu v mezinárodním pohledu a především ve sčítání 2011 v Česku (Šanda, 2022). Článek mimo jiné upozorňoval na fakt, že v současném období dynamických proměn sčítání lidu je pro interpretaci výsledků podstatné mít povědomí o celkovém konceptu sčítání a zejména o některých metodických postupech jeho zpracování. Po zpracování a publikování prvních výsledků sčítání 2021 se toto tvrzení potvrdilo.

Výsledný počet obyvatel podle sčítání 2021 mnohé uživatele dat překvapil. Počet obvykle bydlících obyvatel byl 10 524 167 a počet registrovaných obyvatel (nepřesně označovaných jako „trvale bydlící“⁽²⁾) činil 10 487 748. Ve srovnání se sčítáním 2011 tak počet obvykle bydlících vzrostl o 88 tisíc, počet registrovaných obyvatel se téměř nezměnil, resp. slabě poklesl (o necelé 3 tisíce). Přitom podle intercenzální bilance vzrostl mezi sčítáním 2011 a koncem roku 2020 počet registrovaných obyvatel o 211 168 na 10 701 777. Sčítáním 2021 a zpětnou demografickou bilancí od rozhodného okamžiku k 1. 1. 2021 byl tento stav korigován na 10 494 836.

1) U Český statistický úřad, kontakt: robert.sanda@czso.cz.

2) „Trvale bydlící obyvatelstvo“ je zažitý termín. Vedle občanů ČR a cizinců s trvalým pobytem sem však patří také občané členských zemí EU a EFTA (European Free Trade Association) a jejich rodinných příslušníků s přechodným pobytem, občané třetích zemí s dlouhodobým pobytem (vízum nad 90 dní nebo povolení k dlouhodobému pobytu), a cizinci s platným azylem v ČR. Termín registrované obyvatelstvo je proto vhodnější, i když také není bezproblémový.

Skok v časové řadě počtů obyvatel je po novém sčítání běžný, ale vždy se pohyboval v řádu několika desítek tisíc (s výjimkou stotisícového poklesu při sčítání 1970, což však bylo ve specifické situaci po událostech let 1968 a 1969). Dopad sčítání 2021 na oficiální počty obyvatel byl v historii tzv. moderních sčítání bezprecedentní. Ovlivněny tím byly nejen časové řady stavů obyvatelstva, ale rovněž řady dalších demografických ukazatelů, sociálních statistik atd. Přirozeně to vyvolalo v řadách uživatelů pochybnosti o věrohodnosti výsledků. Proto je nezbytné představit postup vymezení obyvatelstva v tomto sčítání podrobněji, než bývalo obvyklé v minulosti.

Urcitý teoretický rámec a ucelenější přehled o vývoji a současnosti situaci populačních cenů poskytují zmíněný článek z roku 2022, v tomto navazujícím textu se proto již zaměříme výhradně na české sčítání 2021.

2. ADMINISTRATIVNÍ ZDROJE DAT O OBYVATELSTVU VE SČÍTÁNÍ 2021

V období mezi sčítáními 2011 a 2021 se udály některé významné změny v okolnostech souvisejících s využíváním administrativních dat pro statistické účely. Tou zřejmě nejdůležitější bylo přijetí nového nařízení EP a Rady 2015/759 v roce 2015 (*Evropská unie*, 2015). Nařízení novelizovalo původní nařízení 223/2009 o evropské statistice (*Evropská unie*, 2009). Mimo jiné do něho zavedlo nový článek 17a, upravující přístup národních statistických institucí k administrativním záznamům a jejich využívání pro statistické účely. Podle odstavce 1 tohoto článku mají statistické úřady členských zemí „...*právo na rychlý a bezplatný přístup a využívání všech administrativních záznamů a právo tyto záznamy integrovat do statistik, a to v míře nezbytné pro vývoj, vypracovávání a šíření evropské statistiky určené v evropském statistickém programu...*“. Administrativními záznamy jsou (podle odstavce 3 článku 17a) míněny záznamy vedené v systémech veřejné správy.

To poskytlo Českému statistickému úřadu oporu při vznášení požadavků na zpřístupnění administrativních dat.

Důležitým aspektem byl také posun v ochotě ústředních orgánů sdílet data ze svých zdrojů, pokud je takové sdílení v souladu s právním rádem. Zatímco pro sčítání 2011 byla k dispozici pouze jedna tabulka s výčtem osob a jejich základních charakteristik z evidence obyvatel, pro sčítání 2021 bylo k dispozici výrazně více zdrojů. Ve srovnání s řadou jiných států byl však obsah administrativních údajů stále útlý.

Pro sčítání 2021 byly využity následující administrativní zdroje o osobách:

Základní registr obyvatel a agendové systémy vedené Ministerstvem vnitra

Základní registr obyvatel (ROB) poskytoval pro účely sčítání stěžejní údaje (identifikační údaje, adresu pobytu, datum narození, občanství atd.) o státních občanech České republiky, registrovaných občanech členských států EU a EFTA (resp. občanech států, které jsou vázány mezinárodní smlouvou sjednanou s Evropským společenstvím, a jejich rodinných příslušníků), cizincích s povoleným dlouhodobým nebo přechodným pobytem, cizincích s uděleným azylem či doplňkovou ochranou.³⁾

Do základního registru obyvatel se údaje zapisují výhradně automaticky, prostřednictvím několika tzv. agendových informačních systémů. Agendovým systémem evidujícím české občany je systém evidence obyvatel (AISEO), který vznikl z původního informačního systému evidence obyvatel (ISEO) – ten původně evidoval občany i cizince a byl jediným administrativním zdrojem dat o osobách při sčítání 2011. Údaje o cizincích jsou do ROB zapisovány prostřednictvím agendového informačního systému cizinců (AISC). Tyto agendové systémy obsahovaly některé rozšiřující informace nad rámec ROB (například pohlaví, rodinný stav, počátek pobytu nebo základní rodinné vazby mezi osobami).^{4,5)}

3) Vymezení registrovaných osob a okruh evidovaných údajů v základním registru obyvatel přesně definuje zákon 111/2009 Sb., o základních registrech (*Česko*, 2009).

4) Novelou zákona o základních registrech účinnou od února 2022 se ROB rozšířil o údaje o pohlaví a rodinném stavu, v době sčítání 2021 však byly tyto charakteristiky dostupné pouze v agendových systémech.

5) Dalšími systémy editujícími údaje v ROB jsou agendový systém evidence občanských průkazů, agendový systém evidence cestovních dokladů a agendový systém evidence datových schránek. Ty však pro sčítání přímo využity nebyly, pouze jejich údaje propané do ROB.

Centrální registr pojištěnců

Centrální registr pojištěnců (CRP) sdružuje základní informace o účastnících zdravotního pojištění. Vede ho Všeobecná zdravotní pojišťovna České republiky a pro sčítání byl vedle registru obyvatel jediným administrativním zdrojem pokrývajícím téměř veškeré obyvatelstvo. Neobsahoval sice žádné údaje, které by samy o sobě byly využitelné pro pokrytí některých témat zjišťovaných sčítáním, ale poskytl velmi cenné informace o přerušení zdravotního pojištění z důvodu pobytu občana v zahraničí. Vzhledem k tomu, že se podstatná část osob stěhujících se dlouhodobě do zahraničí neodhláší z pobytu na území ČR, byly údaje v CRP o přerušeném pojištění velmi cenné při identifikování osob vedených v registru obyvatel, které ale neodpovídaly definici obyvatele Česka.

Subsystémy informačního systému sociálního zabezpečení

Integrovaný informační systém sociálního zabezpečení (IISZ) vede Česká správa sociálního zabezpečení. Pro sčítání 2021 z něho byly využity tři subsystémy. Prvním byl subsystém důchodů, z něhož měl ČSÚ pro sčítání 2021 k dispozici seznam osob, které k rozhodnému okamžiku sčítání pobíraly starobní nebo invalidní důchod. Druhým byl subsystém peněžité pomoci v mateřství – seznam osob přijímajících peněžitou pomoc v mateřství. Posledním je subsystém pojistných vztahů, z něhož lze zjistit zaměstnání osob (v rozdělení na postavení zaměstnanec/podnikatel), včetně IČO zaměstnavatele.

Databáze regionálního školství a sdružené informace matrik studentů

Správce těchto zdrojů je Ministerstvo školství, mládeže a tělovýchovy. Záznamy z regionálních školských matrik obsahovaly údaje za rok 2021 o žácích základních, středních a vyšších odborných škol a konzervatoří. Ze sdružených informací matrik studentů byla k dispozici data o vysokoškolských studiích od roku 1999. Pro sčítání byly tyto zdroje využitelné zejména jako jeden ze zdrojů pro zařazení osob do kategorií ekonomické aktivity (žáci, studenti), zároveň lze údaje v omezené míře využít pro získání dat o dosavadní úrovni dosaženého vzdělání studentů.

Systémy Ministerstva práce a sociálních věcí (MPSV)

Z MPSV byly pro sčítání k dispozici záznamy o uchazečích a zájemcích o zaměstnání (systém OK práce), o příjemcích pomoci v hmotné nouzi (systém OK nouze) a o příjemcích rodičovského příspěvku a jejich dětí (systém OK registr).

Údaje z přiznání k dani z příjmů fyzických osob

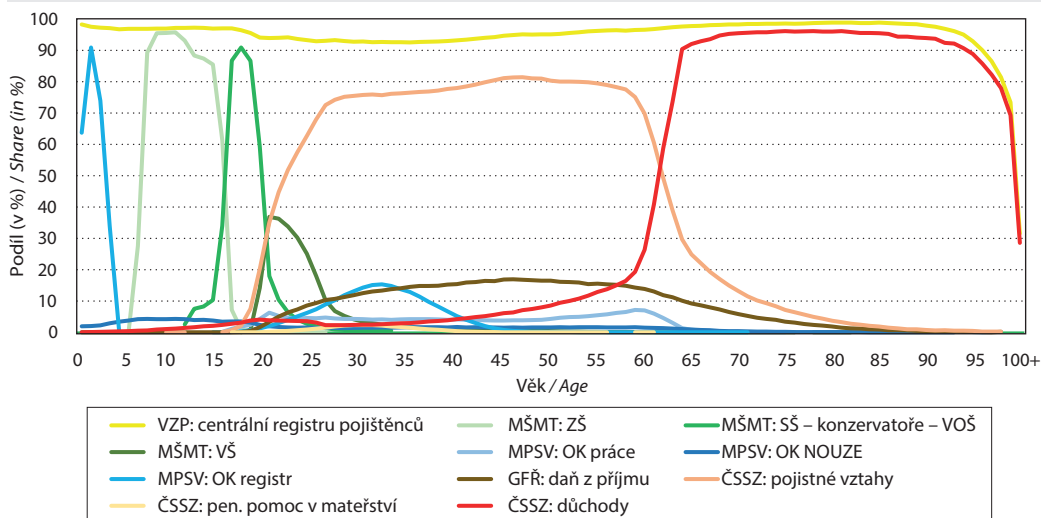
Údaje poskytuje Generální finanční ředitelství ČSÚ primárně pro účely jiných statistik, právní předpisy ovšem umožňují i využití pro sčítání lidu. Obsahují informace o ekonomické činnosti osob podávajících přiznání k dani z příjmů. Vzhledem k tomu, že data byla k dispozici s velkým časovým odstupem od období, k němuž se vztahovala, pro sčítání 2021 připadalo v úvahu pouze využití dat z přiznání za rok 2020. Využití těchto dat pro sčítání bylo tudíž spíše okrajové.

Jak již bylo zmíněno, pouze ROB a CRP jsou administrativními zdroji pokrývajícími víceméně celou populaci. Ostatní zdroje obsahují údaje o specifických skupinách obyvatel (nezaměstnaní, příjemci důchodů, žáci, studenti, ...), představovaly tedy spíše určitou „mozaiku“, kterou bylo třeba složit, aby bylo podchyceno obyvatelstvo jako celek. To je patrné z obr. 1, který znázorňuje podíly osob evidovaných v registru obyvatel nalezených v daném administrativním zdroji. Děti do 3–4 let vede hlavně systém OK-registr evidující rodičovské příspěvky, děti a mladé lidi do věku zhruba 18 let jednotlivé databáze regionálního školství, dospělou populaci v produktivním věku zachycuje (vedle zdravotního pojištění) z větší části pouze systém ČSSZ-pojistné vztahy spolu s daty GFŘ z daňových přiznání, seniorskou populaci zejména systém ČSSZ-důchody. V souhrnu tyto zdroje dat neposkytovaly nad rámec ROB-AISEO-AISC příliš údajů, aby bylo možné zredukovat obsah sčítacích formulářů. Ve sčítání 2021 se výhradně z ostatních zdrojů o obyvatelstvu získával pouze údaj o postavení v zaměstnání (v členění na zaměstnanec/podnikatele bez zaměstnanců a podnikatele se zaměstnanci).

Z hlediska charakteristik, které pro sčítání poskytují, byl tak zdaleka nejvyužitelnější registr obyvatel, resp. soustava ROB-AISEO-AISC.

Obr. 1: Podíly osob s registrovaným pobytem v ROB nalezených ve vybraných dalších zdrojích podle věku

Shares of persons with registered residence in the population register (ROB) covered in selected supplementary sources by age



Zdroj: ČSÚ, 2020.

Source: ČSÚ, 2020.

Pozn./Note: VZP: centrální registr pojištěnců – The General Health Insurance Co.: The central health insurance register; MŠMT: ZŠ – The Ministry of Education, Youth and Sports (MŠMT): grammar schools; MŠMT: SŠ – konzervatoře – VOŠ; MŠMT: secondary, post-secondary schools, conservatoire, MŠMT: VŠ – MŠMT: tertiary; MPSV: OK PRÁCE – The Ministry of Labour and Social Affairs (MPSV) – The register of unemployment; MPSV: OK NOUZE – MPSV: The system of assistance in material need; MPSV: OK REGISTR: The Social supports register; GFR: daň z příjmu – The General Finance Directorate: The income tax register; ČSSZ: pojistné vztahy – The Social Security Administration (ČSSZ): The social security register; ČSSZ: pen. pomoc v mateřství – ČSSZ: The maternity beneficiary register; ČSSZ: důchody – ČSSZ: The pension beneficiary register.

3. VYMEZENÍ OBYVATELSTVA ČESKA VE SČÍTÁNÍ LIDU 2021

Při tradičním sčítání je zařazování sečtených osob do obyvatelstva relativně přímočaré. Prostřednictvím dotazníků se získají údaje o co největším počtu osob, které potenciálně patří do definované populace (přítomné/trvale bydlící/obvykle bydlící...), a vyhodnotí se získané odpovědi. Například při posledním ryze tradičním sčítání u nás, sčítání 2001, byly do obyvatelstva zahrnovány osoby, které měly na území státu hlášen trvalý nebo dlouhodobý pobyt. Sčítací komisaři distribuovali mezi obyvatelstvo sčítací formuláře, vyplněné formuláře poté sesbírali. Následně se zpracovávají údaje o adresách uvedených na sčítacích listech osoby a odpovědi na otázku *druh pobytu*.

Rodná čísla byla využita k odhalení vícekrát sečtených osob a vyřazení duplicit. Počet shromážděných formulářů, úplnost a kvalita vyplnění adres a rodných čísel určovaly výslednou velikost populace (podle závěrečné zprávy pro vládu výsledky podhodnocovaly skutečný stav populace asi o jedno procento: ČSÚ, 2004).

Při kombinovaném sčítání, využívajícím celoplošné dotazníkové šetření a k němu dostupná administrativní data, je problematika vymezení populace složitější. Postup uplatněný v prvním českém kombinovaném cenzu, konaném v roce 2011, popisuje např. ČSÚ (2013) nebo Šanda (2015). Sčítání v roce 2021 sice využívalo více administrativních zdrojů než sčítání 2011, ale konstitutivní zdroje⁶⁾ byly opět dva – dotazníkové šetření a soustava ROB-AISEO-AISC.

6) Termínem konstitutivní zdroj je míněn takový zdroj, který „přidává“ do statistického výsledného souboru záznamy (v popisovaném případě osoby). Ostatní zdroje jsou označovány jako doplňkové – jsou to zdroje, které obsahují relevantní údaje rozšiřující informace v konstitutivních zdrojích, ale nemohou „přidat“ osobu (může však být na jejich základě osoba ze statistického souboru vyřazena).

3.1 PROPOJENÍ SČÍTACÍCH FORMULÁŘŮ SE ZÁZNAMY Z ROB VE SČÍTÁNÍ 2021

Mezi nejdůležitější kroky zpracování dat při kombinovaných sčítáních patří propojování záznamů z formulářů s administrativními záznamy a analýza „administrativních známek života“ (signs-of-life analýza, blíže Šanda, 2022). Při sčítání 2021 měli respondenti pro účely propojení dat ze sčítacích formulářů se záznamy ROB uvést na formulář jméno, příjmení a buď rodné číslo, nebo kombinaci data narození a typu a čísla osobního dokladu (český občanský průkaz, český pas, doklad vydaný cizinci Českou republikou). V identifikačních údajích osob na sčítacích formulářích se však vyskytovaly chyby v podobě nepřesností v rodných číslech a číslech dokladů, ve jménech a příjmeních uvedených v jiném tvaru než ve tvaru zaznamenaném v ROB. Objevovaly se chyby jako např. uvádění stejného čísla dokladu všem členům domácnosti sečteným na formuláři (překvapivě často spolu s tím bylo uváděno i stejné datum narození) atd. I po provedené standardizaci jmen a příjmení (odstraňování titulů, nealfabetických znaků atd.) zhruba každá

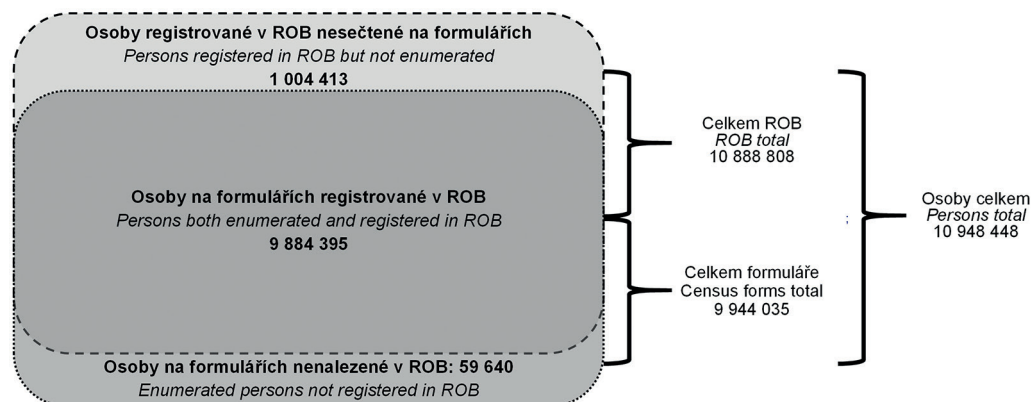
pátá osoba na formuláři neměla identifikační údaje v podobě shodné se záznamem v populačním registru.

Na základě analýzy stavu identifikačních údajů na formulářích a v ROB byla stanovena sada pravidel kombinujících deterministické a pravděpodobnostní postupy propojování těchto dvou zdrojů (mj. Fellegi – Sunter, 1969; Dusetzina et al., 2014). V rámci propojování byly také odhaleny řádově statisíce duplicitně sečtených osob.⁷⁾ V těchto případech byl vybrán prioritní formulář, opět podle série stanovených pravidel, jejichž cílem bylo vybrat nejkvalitnější, resp. nejméně nejvíce duplicitně sečtený formulář pro danou osobu. Výsledek procesu propojování a deduplikace (vyřazení duplicitních formulářů) je znázorněn schématem na obr. 2.

Ve srovnání se sčítáním 2011 se ve sčítání 2021 prohloubil rozdíl mezi počty osob sečtených na sčítacích formulářích a počty registrovaných osob s platným pobytem k rozhodnému okamžiku. Zatímco v roce 2011 byl rozdíl mezi těmito počty 566 tisíc (10 149 tisíc osob na formulářích vs. 10 715 tisíc evidovaných osob s platným pobytem v ISEO), v roce 2021 byl tento rozdíl 945 tisíc osob. Pokles počtu osob na formu-

Obr. 2: Záznamy osob ve zpracování SLDB 2021 podle zdroje dat (výsledek procesu propojení záznamů a vyřazení duplicit)

Personal records processed in the 2021 Population and Housing Census by data source (results of the record linkage and deduplication process)



Zdroj: Cenzový informační systém.

Source: Census information system.

7) Z toho je patrné, jak důležitou roli hraje práce s identifikačními údaji osob – vliv metody a nastavených parametrů pro propojení, resp. určení duplicity na počet obyvatel i na celorepublikové úrovni může být dosti výrazný.

lářích o zhruba 200 tisíc oproti předchozímu cenzu (a cca o 300 tisíc oproti sčítání 2001) potvrzuje obecně pozorovaný postupný pokles účasti obyvatel na sčítání, a to i přes redukcí počtu otázek a značné zjednodušení přístupu k formulářům ve srovnání s minulými cenzy díky zavedení online formulářů.

3.2 VYMEZENÍ OBYVATEL Z OSOB SEČTENÝCH NA FORMULÁŘÍCH

Celkem 59 640 osob, které byly sečteny na sčítacích formulářích, nebylo nalezeno v ROB. Měli by to být především cizinci, kteří na našem území fakticky pobývají, ale nejsou přihlášení k pobytu. V této skupině byly proto za obyvatele považovány pouze osoby, které podle vyplněných údajů nejenže splňovaly kritéria pro zařazení do obvykle bydlícího obyvatelstva, ale zároveň deklarovaly první pobyt po narození v zahraničí (jako náhrada za občanství, které se na formulářích nezjišťovalo). Další podmínkou byly úplné identifikační údaje, bez zjevných chyb. Ostatní formuláře v této skupině byly z dalšího procesu zpracování vyřazeny – byly mezi nimi formuláře s neúplnými či chybnými identifikátory, což bylo důsledkem jejich nepropojení s ROB (a byly by tedy duplicitně započteny do obyvatelstva), nebo fiktivní osoby.⁸⁾ Z uvedených 59 640 osob tak bylo do obvykle bydlícího obyvatelstva zahrnuto pouze 36 tisíc (do registrovaného obyvatelstva nebyl z této skupiny zahrnut nikdo, neboť nikdo z nich neměl registrovaný pobyt). Tento přístup byl striktnější než při sčítání 2011, při němž byly za obvykle bydlící obyvatele považovány všechny osoby z formulářů, pokud vyplněné údaje odpovídaly definici obvykle bydlícího obyvatelstva.

Celkem 9 884 395 osob bylo sečteno na formulářích a zároveň mělo registrovaný pobyt v ROB (viz obr. 2). Drtivá většina z nich (9 785 846) na formuláři deklarovala obvyklý pobyt v Česku a byla tak zařazena do obyvatelstva. Zbýlých necelých sto tisíc (přesně 98 549) osob však na formuláři deklarovalo obvyklý pobyt v zahraničí. Ty nebyly zahrnuty do obvykle bydlícího, ale ani do registrovaného obyvatelstva Česka. V roce 2011 tvořilo tuto skupinu 36 tisíc osob, které tehdy (samozřejmě) rovněž nebyly zahrnuty

do obvykle bydlícího obyvatelstva, ale do „trvale bydlícího“ započteny byly. Při zachování přístupu z roku 2011 by tak počet registrovaných obyvatel, kteří následně vstupují do všech intercenzálních počtů průběžné statistiky obyvatelstva, byl o téměř sto tisíc vyšší, dvoustetisícový propad počtů obyvatel mezi roky 2020 a 2021 by se tak zmínil na polovinu. Z pohledu celkového přístupu ke zpracování, jehož cílem bylo zahrnout do populace osoby reálně žijící na našem území, se ale přístup z roku 2011 jeví jako nelogický.

3.3 VYMEZENÍ OBYVATEL Z REGISTROVANÝCH OSOB NESEČTENÝCH NA FORMULÁŘÍCH (SIGNS-OF-LIFE ANALÝZA)

Registrovaných, ale na formulářích nesečtených osob bylo 1 004 413 (ve sčítání 2011 to bylo 693 tisíc). Záznamy v této skupině se mohly týkat osob, které na našem území v rozhodný okamžik žily, ale sčítání se nezúčastnily, a osob žijících v zahraničí, které oficiálně neohlásily ukončení pobytu. Mohli to být také zemřelí, jejichž úmrtí nebylo v registru zaevidováno. Vybrat z této skupiny osoby, které pravděpodobně fakticky („obvykle“) na našem území k rozhodnému okamžiku žily, bylo cílem signs-of-life analýzy.

V roce 2011, kdy vedle ISEO nebyl pro účely sčítání k dispozici žádný administrativní zdroj dat o osobách, byly možnosti výběru nesečtených osob evidovaných v ISEO velmi omezené. Dostupné zdroje pro sčítání 2021 sice rovněž v mezinárodním srovnání nepředstavovaly bohatou datovou základnu, ale umožnily již provést relativně ucelenou analýzu. Konkrétní pravidla se odvíjela zejména od zjištění, jak se v administrativních zdrojích „projevovali“ respondenti, kteří se prostřednictvím formuláře sečetli, přičemž deklarovali obvyklý pobyt v Česku a zároveň byli evidováni v ROB. Z vyhodnocení této skupiny obyvatel mimo jiné vyplývalo, že občané ČR měli v celém věkovém spektru (až na výjimky v řádu nízkých jednotek procent) platné zdravotní pojištění v centrálním registru pojištěnců. Zároveň byli, kromě dětí ve věku 0 let a 3 až 6 let, vedeni ještě v dalším doplňkovém admi-

8) Z individuální kontroly vybraných formulářů byly evidentně fiktivní osoby na formulářích nalezeny.

nistrativním zdroji. Děti ve věku 1–2 roky byly spolu s jejich rodiči vedeny v systému OK registr evidujícím příjemce rodičovského příspěvku (3leté děti už tento zdroj pokrýval pouze zhruba z 60 %), občané ve věku 70 a více let byli navíc téměř vždy příjemci důchodu atp.

Na základě těchto zjištění, konzultací s pracovníky rezortů spravujících jednotlivé zdroje a dalších odborných debat byla stanovena následující pravidla, z nichž alespoň jedno muselo být splněno, aby byla daná osoba zahrnuta do obyvatelstva:

- 1) Osoba byla k rozhodnému okamžiku evidována v systému OK práce (zájemci a uchazeči o zaměstnání) nebo v systému OK nouze (příjemci pomoci v hmotné nouzi) nebo v IISZ jako příjemce peněžitě pomoci v mateřství. Tyto informační systémy evidují osoby po relativně krátkou dobu a vzhledem k jejich charakteru je velmi pravděpodobné, že evidovaná osoba se na území Česka skutečně nachází.
- 2) Osoba byla evidována v centrálním registru pojištěnců jako pojištěnec s platným zdravotním pojištěním a zároveň
 - a) byla dítětem ve věku 0–6 let a některý z jejích rodičů, případně osoba oprávněná pobírat na dítě rodičovský příspěvek byli sečtení či vykazovali administrativní „známky života“
 - b) nebo byla ve věku 7–15 let a byla evidována v databázích regionálního školství
 - c) nebo byla ve věku 16–69 let a byla evidována v kterémkoliv z dalších zdrojů (tj. byla zároveň v ROB, centrálním registru pojištěnců a ještě v některém z dalších zdrojů)
 - d) nebo byla ve věku 70 a více let a byla poživatelkou starobního důchodu
 - e) nebo u ní v posledních 12 měsících došlo k aktualizaci záznamu v centrálním registru pojištěnců nebo k některé ze specifikovaných změn v ROB resp. AISEO nebo AISC (sňatek na území Česka, narození dítěte na území Česka, přestěhování)
 - f) nebo byla ve věku 17–69 let a v centrálním registru pojištěnců byla vedena jako pojištěnec, za kterého neplatí pojištění stát. Toto pravidlo bylo poměrně benevolentní a splnila ho velká většina osob, která splnila některé z předchozích pravidel. Přesto předchozí pravidla nejsou

zbytečná, protože překryv není stoprocentní a jsou silnějším potvrzením, že daná osoba patří do obyvatelstva.

3. Pokud se jednalo o cizince, změny v posledním roce uvedené v pravidle 2e byly dostatečnou známkou života bez ohledu na to, zda byl cizinec nalezen v centrálním registru pojištěnců.

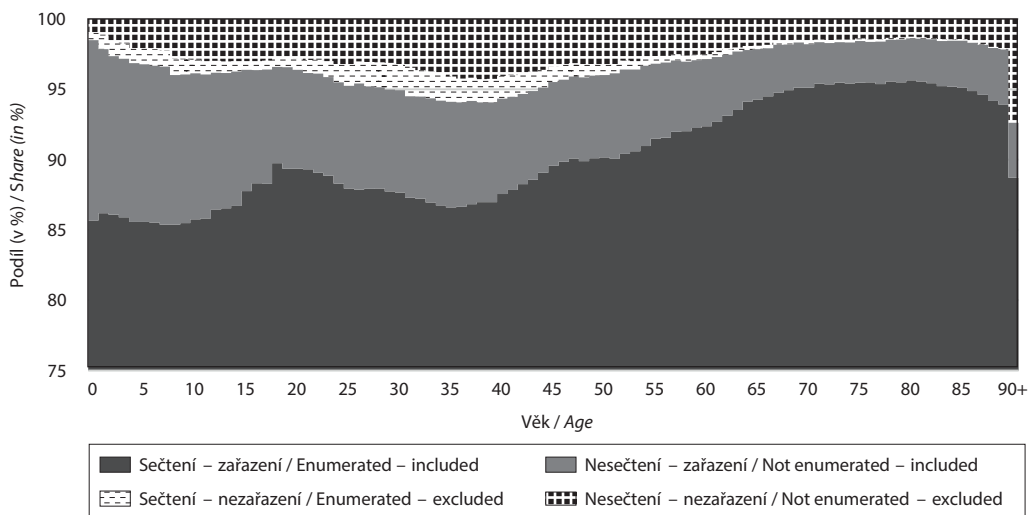
Aplikováním uvedených pravidel bylo identifikováno 702 214 osob evidovaných v ROB s platným pobytem v ČR, které měly dostatečné administrativní známky života, tj. splnili alespoň jednu z uvedených podmínek. Tyto osoby byly zařazeny do obyvatelstva. Zbýlých 302 199 evidovaných osob administrativní známky života nevykazovalo, z hlediska sčítání se tedy jednalo o skupinu představující tzv. přesah pokrytí (*overcoverage*, tedy osoby evidované v registru obyvatel, které ve skutečnosti nepatří do obyvatelstva). Přičteme-li k nim výše zmíněnou skupinu 98 549 osob, které na formuláři deklarovaly obvyklý pobyt v zahraničí, výsledný přesah pokrytí v ROB činil 3,7 %, což byla ve srovnání s obdobným údajem reportovaným jinými zeměmi poměrně vysoká hodnota (více o tomto tématu Šanda, 2022). Přesah pokrytí je důležité téma, a to z pohledu registrace obyvatel jako takové (indikuje míru souladu administrativních dat s „reálným“ stavem), ale i přímo z pohledu populační statistiky. V důsledku anonymizace výsledků sčítání totiž později nelze zjistit, zda např. zemřelá osoba reportovaná ČSÚ matričním úřadem do obyvatelstva byla či nebyla zahrnuta. Vyšší přesah pokrytí znamená i vyšší riziko, že do pohybu obyvatelstva bude zahrnuta demografická událost, která se vymezeného obyvatelstva netýká. Jedním z rizik je také duplicitní sečtení ve více státech (vystěhovalá osoba neodhlášená z pobytu je pravděpodobně registrována též ve své aktuální zemi pobytu, což může např. přispívat k umělému navyšování obyvatelstva Evropské unie).

4. VÝSLEDEK VYMEZOVÁNÍ OBYVATELSTVA

Obr. 3 shrnuje výsledek procesu zařazování registrovaných osob do obyvatelstva v závislosti na věku. Dotazníkovým šetřením v terénu se podařilo nejlépe podchytit osoby ve vyšších věkových skupinách, od 65, resp. 70 let. Ti senioři, kteří se na formuláři nesečetli, byli relativně výrazně dopl-

Obr. 3: Struktura osob registrovaných v ROB podle věku, zdroje dat a zařazení/nezařazení do obyvatelstva ve sčítání lidu 2021

Persons registered in ROB by age, data source, and inclusion/non-inclusion in the population in the 2021 Population and Housing Census



Zdroj: ČSÚ, 2020.
Source: ČSÚ, 2020.

nění do obyvatelstva na základě výsledku signs-of-life analýzy, vyřazena byla jen menší část. Výjimkou byly osoby v nejvyšším věku (nad 90 let), což je jistě z větší části projev neregistrovaných úmrtí.

Nejméně zachycovaly sčítací formuláře dětskou populaci, což bylo poměrně překvapivé (i ve sčítání 2011 se relativně hodně dětí doplnilo z tehdejšího ISEO, ale nepatřily mezi skupiny nejhůře zachycené v terénu). Výrazná většina z nesečtených dětí byla signs-of-life analýzou do obyvatelstva zařazena. Lze spekulovat, že je to důsledek minimálního kontaktu respondentů s komisaři, resp. s pracovníky kontaktních míst. Na rozdíl od všech dosavadních cenzů většina respondentů sčítání 2021 nepřišla do kontaktu s nikým, ale pouze vyplnila online sčítací formulář, což mohlo ve větší míře vést k opomenutí skutečnosti, že i na děti se vztahuje povinnost sčítání.

Další skupinou registrovaných osob, u nichž byl podíl sečtení na formulářích poměrně nízký, byly osoby ve věkové kategorii 30–39 let. I v těchto věcích byla většina nesečtených osob z ROB doplněna do obyvatelstva, ale počty vyřazených osob z obyvatelstva zde byly relativně vysoké. Mezi hlavní příčiny mohly patřit relativně vysoké zastoupení cizinců

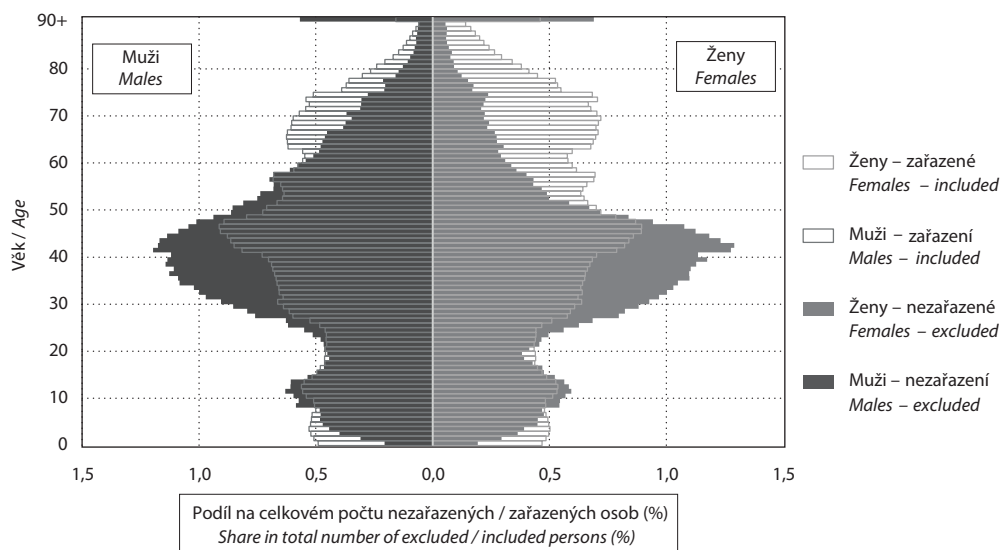
v tomto věkovém intervalu, vyšší intenzita migrace a asi i obecně nižší ochota účastnit se sčítání ve srovnání se staršími generacemi (pro vysvětlení by ale byla nutná výrazně podrobnější analýza a možná i specifické šetření).

4.1 REGISTROVANÉ OSOBY V ROB VYŘAZENÉ ZE SČÍTÁNÍ

Základní rysy věkové struktury vyřazených osob zhruba korespondovaly s poznatky z roku 2011. V roce 2021 byly mezi vyřazenými relativně výrazně (ve srovnání se sečtenou populací) zastoupeny osoby ve věku 30–45 let, u mužů až do 60 let věku (v roce 2011 to bylo 25–40 let, u mužů do 55), a zejména osoby v nejvyšších věcích (v ROB byly registrovány i desítky osob starších 120 let). Stejně jako v roce 2011 byli mezi vyřazenými častěji zastoupeni muži, ovšem oproti roku 2011 byla převaha mužů zřetelná až zhruba od 50 let věku. U mladších osob nebyl vliv pohlaví podstatný, dokonce ve věku mezi 25–30 roky a kolem 40. roku věku ženy převažovaly. Zatímco tedy v roce 2011 se problém přesahu pokrytí v ROB týkal především mužů, v roce 2021 to již platilo

Obr. 4: Osoby registrované v ROB podle věku, pohlaví a zařazení/nezařazení do obyvatelstva ve sčítání lidu 2021

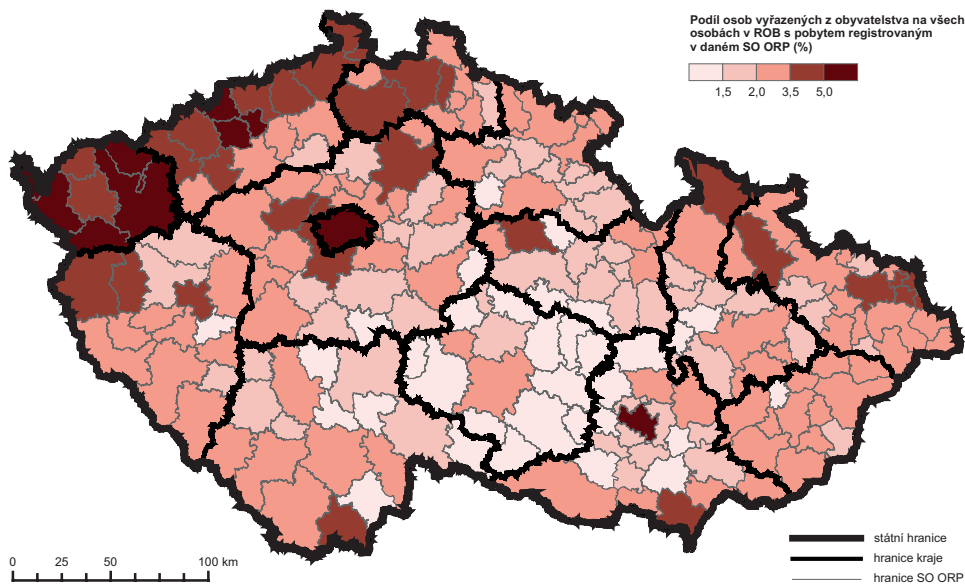
Persons registered in ROB by age, sex and inclusion/non-inclusion in the population in the 2021 Population and Housing Census



Zdroj: ČSÚ, 2020.
Source: ČSÚ, 2020.

Obr. 5: Osoby registrované v ROB vyřazené z obyvatelstva SLDB 2021 podle SO ORP registrovaného pobytu

Persons registered in ROB excluded from the population in the 2021 Population and Housing Census by the microregion (SO ORP) of registered residence



Zdroj: ČSÚ, 2020.
Source: ČSÚ, 2020.

jen omezeně (mezi vystěhovalými občany Česka do zahraničí ženy v současné době převažují, je to tedy logické vyústění neúplné evidence emigrace). Pohlavně věkovou strukturu vyřazených osob znázorňuje obr. 4. Obdobný obrázek za rok 2011 je v článku z roku 2022⁹⁾ – z porovnání je patrný nejen zmíněný růst podílu žen, ale i určité stárnutí vyřazených osob. Tomu zřejmě přispívají oso-

by vystěhovalé do zahraničí již před delším časem, které dlouhodobě zůstávají registrované k pobytu u nás.

Stejně jako v roce 2011 byl pro vyřazené osoby charakteristický poměrně vysoký podíl cizinců. Tab. 1 obsahuje přehled sečtenosti a výsledného zařazení registrovaných osob podle občanství. Ze všech obyvatel v ROB s registrovaným pobytem na našem úze-

Tab. 1: Osoby registrované v ROB podle vybraných občanství, zdroje dat a zařazení/nezařazení do obyvatelstva ve sčítání lidu 2021

Persons registered in ROB by selected state citizenship, data source and inclusion/non-inclusion in the population in the 2021 Population and Housing Census

Státní občanství <i>Citizenship</i>	Registrovaní v ROB <i>Registered in ROB</i>	Registrovaní -sečtení <i>Registered- enumerated</i>		Registrovaní -nesečtení <i>Registered-not enumerated</i>		Podíl (%) / Share of (%)		
		zařazení <i>included</i>	nezařazení <i>excluded</i>	zařazení <i>included</i>	nezařazení <i>excluded</i>	sečtených na registrovaných <i>enumerated in in the registered population</i>	nesečtených-zařazených na obyvatelstvu not enumerated-included in population	nezařazených na registrovaných <i>excluded from the population in registered persons</i>
Osoby v ROB celkem <i>Persons in ROB total</i>	10 888 808	9 785 534	98 861	702 214	302 199	90,8	6,7	3,7
Občané ČR <i>Czech citizens</i>	10 249 602	9 393 679	87 856	603 452	164 615	92,5	6,0	2,5
Cizinci – celkem <i>Foreign nationals</i>	639 206	391 855	11 005	98 762	137 584	63,0	20,1	23,2
Podíl občanů ČR <i>Share of Czech citizens</i>	94,1	96	88,9	85,9	54,5	x	x	x
Německo <i>Germany</i>	20 764	4 217	265	1 331	14 951	21,6	24,0	73,3
Polsko <i>Poland</i>	20 563	11 226	180	3 095	6 062	55,5	21,6	30,4
Rusko <i>Russia</i>	42 229	31 243	708	4 349	5 929	75,7	12,2	15,7
Slovensko <i>Slovakia</i>	123 877	73 148	918	20 775	29 036	59,8	22,1	24,2
Vietnam <i>Vietnam</i>	63 841	48 770	2 160	5 055	7 856	79,8	9,4	15,7
Ukrajina <i>Ukraine</i>	168 554	113 112	2 793	35 679	16 970	68,8	24,00	11,7

Zdroj: ČSÚ, 2020.

Source: ČSÚ, 2020.

9) Šanda, 2022.

mí tvořili cizinci 6 %, k rozhodnému okamžiku jich bylo 639 206. Podle výsledků zpracování však necelá čtvrtina z nich (148 589) na našem území (*obvykle*) nežila. Specifické byly registrované osoby s německým občanstvím, z nichž bylo z obyvatelstva vyloučeno 73 %. Podstatná část z nich byla registrována v Ústeckém kraji (např. ve městě Výsluní představovali vyřazení občané Německa třetinu registrovaných osob, většina z nich byla hlášena na jedné adrese s číslem evidenčním).

Vysokými podíly vyřazených osob se vyznačovala zejména souvislá oblast severozápadního příhraničí (obr. 5). Ve správním obvodu ORP Bílina bylo z obyvatelstva vyřazeno 14 % registrovaných osob, to bylo na úrovni SO ORP nejvíce v republice. Většinu z nich představovali občané Německa hlášení v městě Bílina a v Hrobčicích. Za Bílinou následovalo Brno, kde bylo vyřazeno 8 % registrovaných osob. Absolutně to bylo 32 tisíc, z toho zhruba 20 tisíc bylo registrováno na adrese ohlašovny pobytu (Brno se vyznačovalo celkově nejvyšším podílem osob registrovaných na ohlašovně ze všech SO ORP). Třetí byla v měřítku SO ORP Praha, která je coby region s velmi vysokou intenzitou obratu zahraniční migrace poměrně silně „postížena“ podregistrovanou emigrací. Obcí s nejvyšším podílem vyřazených osob bylo již zmíněné město Výsluní (36 % registrovaných osob vyřazeno). Celkem v 18 obcích představovaly vyřazené osoby více než 10 % registrovaných.

Naopak nízké podíly vyřazených osob byly pozorovány hlavně na Vysočině. Na Vysočině byla celkově vysoká míra pokrytí registrovaných osob terénním šetřením, tudíž ani „nezbylo“ větší množství registrovaných osob k vyřazování nebo doplňování. Vysočina byla i ve sčítání 2011 charakteristická nízkými relativními počty vyřazených osob, nízkými počty obyvatel doplněných z evidence, ale také například poměrně vysokou kvalitou vyplněných sčítacích formulářů (viz např. Šanda, 2020). V měřítku správních obvodů ORP bylo vyřazených relativně nejméně v obvodu

Hlinska (těsně pod 1 %), pod 1,1 % vyřazených bylo ještě v SO ORP Bystřice pod Pernštejnem, na Litovelsku a Třebíčsku. Celkem v 750 obcích nebyl vyřazen nikdo. Vesměs to byly menší obce, většinou pod 150 registrovaných osob, největší byl Zbýšov v okrese Vyškov se 700 registrovanými osobami.

4.2 NESEČTENÉ OSOBY DOPLNĚNÉ DO OBYVATELSTVA SLDB

Pohlavně věková struktura 702 214 osob doplněných do obyvatelstva se ve srovnání s osobami vyřazenými vyznačovala výraznější převahou mužů, tvořili celkově 57 % doplněných obyvatel, ve věkovém rozpětí 27–63 let přes 60 %. Výraznější zastoupení mužů, osob zhruba ve věku mezi 30–40 roky a také cizinců (20 % z výsledné populace cizinců bylo doplněno z ROB – viz tab. 1) byly zhruba očekávané rysy již na základě zkušeností ze sčítání 2011. Již byla řeč o relativně vysokých počtech nesečtených dětí doplněných do obyvatelstva: Děti mladší 15 let představovaly více než čtvrtinu z doplněných obyvatel a na výsledném počtu dětí v populaci se podílely z 11 %.¹⁰⁾

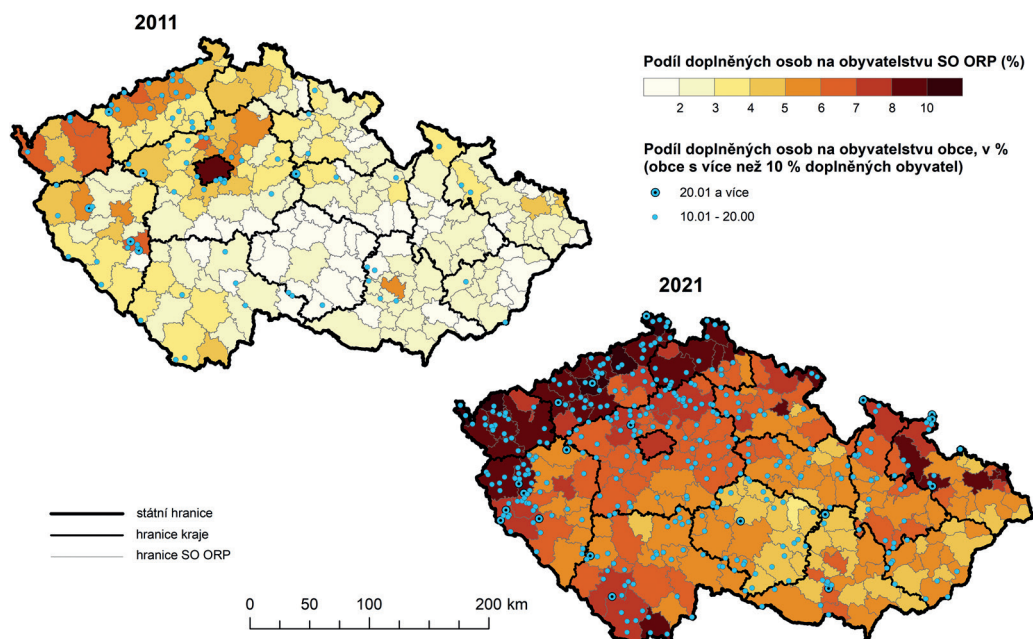
Na obr. 6 je srovnání podílů doplněných osob z ROB (resp. v roce 2011 z ISEO) do obyvatelstva v posledních dvou cenzech (tedy vyšetřenost v terénním šetření). Prostorový vzorec z roku 2011 se ve sčítání 2021 do velké míry zopakoval. Na výsledném počtu obyvatel se nesečtené osoby doplněné z ROB nejvíce projevil v severozápadní polovině Čech, zejména v Karlovenském a Ústeckém kraji, nejméně doplněných osob bylo na Vysočině. Na rozdíl od roku 2011 nepatřila mezi území s nejintenzivněji doplňovaným obyvatelstvem Praha.

Popsané územní rozdíly (na úrovni SO ORP) poměrně dobře odpovídaly územním rozdílům v některých (částečně vzájemně provázaných) charakteristikách zjišťovaných sčítáním. Patrná byla souvislost s religiozitou (s rosoucím podílem věřících klesal podíl doplněných osob), s podílem sečtených osob s prvním bydlištěm po narození v Česku, o něco

10) Děti do 1 roku z 13 % – nejmladší děti byly nejméně sečtenou skupinou již v roce 2011, pro starší děti byl však již podíl doplněných v roce 2021 mnohem větší než v předchozím cenzu, přestože tehdy byl počet doplněných dětí zřejmě příliš vysoký – v roce 2011 byly totiž doplněny všechny evidované děti do 5 let věku, které měly známou adresu registrovaného pobytu (generace 2007 – 2010 tak mohly být ve sčítání 2011 a tím v navazujících intercenzálních stavech mírně nadhodnoceny, čemuž výsledek signs-of-life analýzy 2021 odpovídá).

Obr. 6: Podíl nesečtených obyvatel doplněných z ISEO/ROB na registrovaném („trvale bydlícím“) obyvatelstvu ve sčítáních lidu 2011 a 2021

Shares of not enumerated inhabitants included on the basis of ISEO/ROB in the registered ('permanent resident') population at the microregional level (2011 and 2021 Population and Housing Censuses)



Zdroj: ČSÚ, 2011; ČSÚ, 2020

Source: ČSÚ, 2011; ČSÚ, 2020.

slaběji s podílem rodáků a také s některými socio-ekonomickými charakteristikami, např. mírou nezaměstnanosti. Spojitost mezi těmito ukazateli a úplností, resp. kvalitou výsledků dotazníkového šetření¹¹⁾ naznačovaly i výsledky sčítání 2011 (viz také Šanda, 2020). Tyto pozorované souvislosti naznačují, že socio-ekonomická situace a charakteristiky indukující určitou ukotvenost obyvatel v místě a společnosti, kde žijí, ovlivňují míru participace na sčítání. Vzhledem ke klesající míře této participace u nás i v zahraničí (obecně klesající response v sociálních šetřeních analyzuje např. *National Research Council*, 2013) a dopadům na populační cenzy bude třeba tuto problematiku podrobněji rozpracovat.

Z uvedeného je také zřejmé, že územní rozložení relativních počtů vyřazených a zařazených osob si

v nemalé míře odpovídají. To je zčásti ovlivněno mírou pokrytí registrovaných osob sčítacími formuláři – vysoká míra pokrytí znamená celkově menší počty osob, na něž byla aplikována signs-of-life analýza, tj. podíly vyřazených i zařazených osob byly logicky relativně malé, a naopak. Ve velmi hrubých rysech platilo, že s rostoucím počtem registrovaných, ale nesečtených osob, intenzivněji rostl podíl vyřazených osob než podíl zařazených do obyvatelstva.

Při srovnání sčítání 2011 a 2021 je podstatnější než územní rozdíly výrazný celoplošný nárůst podílů doplněných obyvatel, tedy celoplošné snižování vyšetřovacího prostřednictvím formulářů. Počet obcí, v nichž bylo doplněno více než 10 % obyvatel, stoupl z řádu desítek na stovky (včetně výše zmiňované sedmnáctitisícové Bíliny), zhruba u 20 obcí byla

11) Obdobné rysy jsou patrné nejen u podílu doplněných osob (celková vyšetřovací), ale i u míry vyplnění jednotlivých otázek na sebraných formulářích (tzv. položková vyšetřovací).

doplněna více než pětina jejich obyvatel. To už má na výsledné charakteristiky nepřehlédnutelný dopad. U doplněných osob je například místo registrovaného pobytu ztotožněno s obvyklým, což nemusí odpovídat skutečnosti. Informace zjišťované pouze na formulářích (jako národnost, víra či mateřský jazyk) nejsou za tyto obyvatele zjištěny nikdy, úroveň vzdělání bylo z administrativních zdrojů možno doplnit jen velmi omezeně. Pro doplnění ekonomické aktivity, odvětví ekonomické činnosti a částečně i dojížděky nabízejí dostupné administrativní zdroje více možností, směřování deklarovaných údajů s administrativními v rámci jedné charakteristiky je však vždy diskutabilní, u ekonomické aktivity to platí dvojnásob.

5. ZÁVĚR

Zatímco v roce 2011 české sčítání nakročilo do skupiny kombinovaných cenzů, v roce 2021 se v rámci této kategorie výrazně posunulo ve směru ke sčítáním z registrů. Administrativní zdroje dat měly v 2021 již velmi zásadní roli. Poprvé se některé klíčové ukazatele získávaly výhradně či téměř výhradně z nich (věk, pohlaví, rodinný stav, občanství, postavení v zaměstnání), u jiných podstatně přispěly ke zvýšení vyšetřenosti (počet živě narozených dětí u žen, ekonomická aktivita a odvětví, dojížděka). Především však bylo z registru obyvatel doplněno do výsledného obyvatelstva 702 tisíc osob. Počet osob sečtených na formulářích totiž nedosáhl ani hranice 10 milionů, a to ani přes propracovanou informační kampaň, celkově spíše pozitivní obraz sčítání v médiích a přes nebývalé (v historickém i mezinárodním kontextu) snadný přístup k formulářům a jejich minimalizovanému obsahu. Velmi jasně se tím potvrdil všeobecně pozorovaný trend snižování participace obyvatel na sčítání, bez ohledu na snižování náročnosti a bez ohledu na zákonné povinnosti a potenciální sankce. Je možné, že částečně k nízké vyšetřenosti přispěla i covidová pandemie, která mohla část obyvatel odradit od setkání se sčítacím komisařem, resp. s pracovníky kontaktních míst sčítání.

Míra účasti na sčítání je územně diferencovaná, závisí na věku, pohlaví, občanství, stabilitě obyvatelstva a dalších ukazatelích, což už je známá zkušenost z předchozího sčítání a zčásti i ze sčítání v jiných zemích. Vyšetřenost mezi sčítáními 2011 a 2021 kle-

sala plošně a napříč těmito charakteristikami. Velmi výrazný byl tento pokles zejména u dětí, možná částečně v důsledku redukce činnosti sčítacích komisařů a přesunu celého průběhu sečtení na respondenty za využití online formulářů (což ovšem nijak nezpochybňuje zásadní přínos online sčítání, kterým nepochybně celková vyšetřenost vzrostla a kvalita získaných dat byla nesrovnatelně vyšší než u listinných formulářů, o rychlosti a efektivitě sběru dat nemluvě).

Poprvé byla ve sčítání 2021 aplikována komplexnější signs-of-life analýza, při níž se za využití dostupných informací z administrativních zdrojů rozhodovalo o zařazení či nezařazení více než milionu nesečtených osob s registrovaným pobytem do výsledné populace. Jak bylo řečeno, 702 tisíc z nich bylo do výsledků sčítání zahrnuto, 302 tisíc vyřazeno. Přidáme-li k nim osoby, které na formuláři deklarovaly obvyklý pobyt v zahraničí, výsledkem je přes 401 tisíc osob s registrovaným pobytem v ROB, které (pokud jsou výsledky správné) na území Česka nežily. To je 3,7 % všech osob s registrovaným pobytem (ve sčítání 2011 činil tento *přesah pokrytí* 2,6 %, ovšem vzhledem k omezeným možnostem tehdejšího cenzu třeba toto srovnání brát s rezervou).

Na výsledky sčítání navazuje průběžná demografická bilance obyvatelstva, z níž vycházejí každoroční intercenzální stavy obyvatelstva. Ty jsou široce využívány v mnoha oblastech, mimo jiné jako stěžejní podklad pro rozpočtové určení daní (vezmeme-li v úvahu, že sčítání představuje základ pro bilance na příštích deset let, odvíjí se od jeho výsledků alokace řádově bilionů Kč). Výsledkem popsaného procesu zpracování sčítání 2021 a nasazení výsledků jako nového základu pro bilance byl náhlý „skok“ stavu populace o půlnoci z 31. 12. 2020 na 1. 1. 2021 z 10,7 milionů na 10,5 milionů, přičemž podle údajů v ROB mělo v té době registrovaný pobyt 10,9 milionu osob. Nejen pro samosprávy, jejichž příjmy jsou na těchto datech přímo závislé (a jejichž představitelé zároveň mají přístup k údajům o osobách registrovaných v ROB), ale pro všechny uživatele jsou tyto rozdílné údaje matoucí, na lokální a regionální úrovni k tomu ještě přistupuje problém registrovaného vs. obvyklého pobytu. Hlavním smyslem předkládaného článku proto bylo poskytnout základní představu o postupu zpracování sčítání pro lepší porozumění jeho výsledkům (včetně výsledků navazující demografické bilance).

Z pozorovaného zvětšování rozdílu mezi počty sečtených a registrovaných osob (v roce 2001 do 100 tisíc, v roce 2011 zhruba 700 tisíc, v roce 2021 milion) lze stěží usuzovat, že při příštím sčítání nastane v tomto trendu obrat. Signs-of-life analýza bude pravděpodobně mít ještě větší význam než v roce 2021. Pro její další zkvalitnění a aplikovatelnost na výrazně větší část populace, případně na populaci celou, bude třeba se soustředit zejména na skupiny, které jsou zdroji využitými ve sčítání 2021 méně pokryty. Jsou to například ekonomicky neaktivní osoby v produktivním věku, děti před nástupem do školy, pendleři nebo cizinci. Jedním z hlavních úkolů populační statistiky pro nejbližší roky bude proto získání dat z dalších existujících zdrojů a zřejmě také založení zdrojů nových.

Téma signs-of-life analýzy nabývá na významu i proto, že v příštích letech již nebude aplikována jen v desetiletých sčítáních, ale také v každoročních stavech obyvatelstva. V dosavadní praxi se v bilanci obyvatelstva navazují demografické události jednotlivých, identifikovaných osob, na anonymní agregované počty

obyvatel ze sčítání. Není (a nikdy nebylo) zajištěno, že například matka narozeného dítěte je započtena do obyvatelstva, ale narození je zahrnuto do ukazatelů plodnosti. Podobně nikdy nebyla k dispozici informace o tom, zda zemřelá osoba byla sečtena, ale do bilančního stavu se (měla-li v Česku registrovaný pobyt) započítá a úmrtí je zahrnuto do ukazatelů úmrtnosti. V kontextu aktuálního vývoje, kdy se rozevírají nůžky mezi registrovanými osobami a populací vymezenou sčítáním, je popsáný stav obtížněji akceptovatelný. Východiskem je přechod od dosavadního demografického bilančního k vymezování intercenzálních stavů obyvatelstva z identifikovaných osob registrovaných v ROB aplikováním signs-of-life analýzy. Potom by i u jednotlivých demografických událostí bylo možné určit, zda se týkají osob zahrnutých do stavů obyvatelstva či nikoliv. Založení cenového informačního systému novelou zákona o státní statistické službě (Česko, 1995) bylo nejdůležitějším krokem k tomuto cílovému stavu a v současné době (červen 2023) probíhají práce na přípravě této zásadní transformaci populační statistiky.

Literatura

- Česko. 1995. Zákon č. 89 ze dne 29. dubna 1995 o státní statistické službě. In: Sbíрка zákonu České republiky, 1995, částka 19.
- Česko. 2009. Zákon č. 111 ze dne 27. dubna 2009 o státní statistické službě. In: Sbíрка zákonu České republiky, 2009, částka 33.
- ČSÚ. 2004. Závěrečná zpráva o sčítání lidu, domů a bytů 2001. Zasedání vlády ČR, 30. 6. 2004.
https://www.czso.cz/documents/11308/23212108/vlada_sldb_0604.pdf/51f3e68f-66d6-4891-bdbd-374447dc8b38?version=1.0
- ČSÚ. 2013. Sčítání lidu, domů a bytů - Pramenné dílo - 2011. Praha: Český statistický úřad.
- Dusetzina et al. 2014. *Linking Data for Health Services Research: A Framework and Instructional Guide*. Chapel Hill: AHRQ.
- Evropská unie. 2009. Nařízení Evropského parlamentu a Rady (ES) č. 223/2009 ze dne 11. března 2009 o evropské statistice. In: *Úřední věstník Evropské unie*, 2009.
- Evropská unie. 2015. Nařízení Evropského Parlamentu a Rady (EU) 2015/759 ze dne 29. dubna 2015, kterým se mění nařízení (ES) č. 223/2009 o evropské statistice. In: *Úřední věstník Evropské unie*, 2015.
- Fellegi, I. P. – Sunter, A. B. 1969. *A Theory for Record Linkage*. *Journal of the American Statistical Association*, 40, s. 1183–1210.
<https://doi.org/10.1080/01621459.1969.10501049>.
- National Research Council. 2013. *Nonresponse in Social Science Surveys: A Research Agenda*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/18293>.
- Šanda, R. 2015. Obvyklý a trvalý pobyt ve sčítání lidu 2011. *Demografie*, 57(2), 113–126.
- Šanda, R. 2020. *Výšetřenosť údajů o dojíždě ve sčítání lidu v roce 2011 a jejich rekonstrukce pomocí metody Jaro-Winkler*. *Demografie*, 62(1), 27–42.
- Šanda, R. 2022. Administrativní zdroje dat ve sčítáních lidu se zaměřením na sčítání 2011 v Česku. *Demografie*, 64(2), 106–123.

Zdroje dat

- ČSÚ. 2011. Zpracovatelská databáze SLDB 2011 [databáze].
- ČSÚ. 2020. Cenzový informační systém [databáze].

ROBERT ŠANDA

Je absolventem Geografického ústavu Přírodovědecké fakulty Masarykovy univerzity, kde v současnosti pokračuje v doktorském studiu programu sociální geografie a regionální rozvoj. Působí na Českém statistickém úřadě v odboru statistiky obyvatelstva. Podílel se na přípravě a zpracování sčítání lidu, domů a bytů v roce 2021.

SUMMARY

In 2021 Czechia conducted the second combined census in Czech history. The first one, in 2011, used only one source of administrative data on persons. Ten years later, a significantly larger number of registers and databases on persons were employed. The available administrative data still lacked information on some essential characteristics, so it was inevitable that full field enumeration using census forms still had to be conducted. Personal records in the population register (ROB) and records obtained from the census forms were linked by applying deterministic and probabilistic record linkage techniques. This linking process resulted in the creation of a base list of persons. The persons in the base list were then searched for in the other (supplementary) sources to obtain more information on them.

Population registers usually also contain records of people who emigrated without deregistration and sometimes also records of people who died but the death was not registered. This results in an 'overcoverage' problem. Dealing with overcoverage is a key data processing step for improving the quality of combined or register-based censuses. While population register data are often influenced by overcoverage, traditional full field enumeration suffers from undercoverage as a result of declining public willingness to participate in censuses. The 2021 base list of persons contained more than 1 million persons who were registered in ROB and were not found in the census forms (9% of all persons registered in ROB). This was a joint effect of overcoverage (ROB) and undercoverage (enumeration) problems. The task was to split these persons into an 'over-

coverage group' and an 'undercoverage group'. The former should be excluded from the census population, the latter should be included. A (administrative) signs-of-life analysis was applied to deal with this issue. Persons both registered and enumerated (those who declared Czechia as their usual place of residence) were considered a reference subpopulation and their typical age-specific occurrence in the supplementary sources was analysed separately for Czech citizens and foreign nationals. Based on this analysis a set of criteria were defined, and some additional criteria were formulated on the basis of common sense. The records of every person registered but not enumerated were evaluated against these criteria to determine whether to include the person in the census population or exclude them as not living in Czechia.

As a result, we included 702,000 persons in the census population, i.e. they were considered to be residents who had not participated in the census. The remaining 302,000 persons were excluded for not meeting the signs-of-life criteria. The final total population (usually resident) was 10,524,167. This means that the non-response rate of the field enumeration (undercoverage) was 6.7%. The ROB-overcoverage rate was 3.7% (those 302,000 excluded from the population plus 99,000 people both enumerated and registered who declared their usual place of residence as outside Czechia). Compared to the 2011 census, both the undercoverage rate and the overcoverage rate have increased. These trends are likely to continue in the future, so signs-of-life analysis will become even more crucial for maintaining the quality of future censuses and

of population statistics in general. Getting access to more existing data sources and establishing new ones is the key to an improved signs-of-life approach. Improvement is especially necessary for the coverage of subgroups of the population that are currently

poorly covered by the supplementary sources that were available for the 2021 census, namely preschoolers, economically inactive adults, cross-border commuters, and foreign nationals (especially EU citizens).

STATISTIKA V KAPSE

Nejnovější statistické ukazatele ve vašem mobilu

Regionální údaje na základě vaší polohy

Každý den nové aktuální číslo dne

čsú
mobilní aplikace
k dispozici zdarma

Google Play App Store