



Danilo Dolenc

REGISTRSKI POPIS – METODOLOŠKI IZZIVI, REŠITVE IN KAKOVOST



Registrski popis – metodološki izzivi, rešitve in kakovost

Avtor: Danilo Dolenc

Objavil: Statistični urad Republike Slovenije, Litostrojska cesta 54, SI-1000 Ljubljana

Objavljeno: januarja 2017

Dostopno na: <http://www.stat.si/StatWeb/glavnanavigacija/metode-in-klasifikacije/splosno-o-metodah#raziskovalni-prispevki>

Uporaba in objava podatkov, komentarjev in grafikonov sta dovoljeni le z navedbo vira.

Podatki in stališča, izraženi v prispevku, ne odražajo nujno uradnih statističnih podatkov ali uradnega stališča Statističnega urada Republike Slovenije.

REGISTRSKI POPIS – METODOLOŠKI IZZIVI, REŠITVE IN KAKOVOST

Danilo Dolenc¹

Povzetek

V letu 2015 smo v Sloveniji izpeljali že drugi registrski popis prebivalstva, tako da imamo zdaj na razpolago tudi metodološko povsem primerljive podatke. V tem prispevku bomo prikazali izzive, ki so bili pred nami, ko smo se relativno pozno (v začetku leta 2009) začeli intenzivno pripravljati na prvo izvedbo registrskega popisa v letu 2011. Osnovno izhodišče za registrski popis je bilo podatke v administrativnih virih, zbrane v skladu s koncepti posameznega vira, metodološko ustrezno pretvoriti v statistične podatke, tako da bomo lahko v celoti izpolnili obveznosti države v skladu z uredbo o popisih² (prvič do sedaj je bila izvedba popisa tudi zakonska obveza države) ter istočasno našim uporabnikom ponudili čim več vsebin, ki so se zbirale v preteklih terenskih popisih. Oba registrska popisa smo izpeljali po enaki metodologiji, razlika med obema izvedbama je predvsem v poenostavitvi statističnega procesa zaradi izboljšanja kakovosti vhodnih podatkov, kar velja predvsem za podatke Registra nepremičnin. V letu 2015 pa smo pridobili evropska sredstva in tako smo v letu 2016 prvič do sedaj izvedli terensko raziskovanje, katerega namen je bil ugotoviti kakovost podatkov o prebivalstvu v osnovnem administrativnem viru (Centralni register prebivalstva - CRP) ter posledično ovrednotiti precenjenost oz. podcenjenost podatkov o prebivalstvu v registru ter posledično v demografski statistiki.

Ključne besede:

registrski popis, administrativni viri, metodologija, gospodinjstvo, družina, kazalniki kakovosti

¹ Statistični urad Republike Slovenije, danilo.dolenc@gov.si

² Uredba (ES) št. 763/2008 Evropskega parlamenta in Sveta o popisih prebivalstva in stanovanj – v nadaljevanju Uredba.

REGISTER-BASED CENSUS – METODOLOGICAL CHALLENGES, SOLUTIONS AND QUALITY

Danilo Dolenc¹

Abstract

In Slovenia in 2015 the second register-based population census was conducted and provided us with methodologically comparable data. This paper presents the challenges ahead of us when we – relatively late, i.e. in the beginning of 2009 – started the intensive preparations for the first implementation of the register-based census to be held in 2011. The main starting point for implementing such a census was to methodologically adequately convey the data from the administrative registers, which were gathered in line with the concepts of the individual source, into statistical data, with the aim to fully comply with the EU regulation on censuses² (as this was the first time that conducting a census became legal commitment of a country) and also to offer to data users the widest possible range of the data obtained with previous traditional censuses. Both register-based population censuses were conducted in line with the same methodology; the difference between them was the simplification of the statistical process due to improvement of the input data quality, especially in case of the Real Estate Register. In 2015 Slovenia received EU funding and thus in 2016 for the first time field work was carried out in order to determine the quality of population data from the basic administrative source (Central Population Register, CPR) and to subsequently evaluate the rate of overcoverage and undercoverage in the register and also demography statistics.

Keywords:

register-based census, administrative sources, methodology, household, family, quality indicators

¹ Statistical Office of the Republic of Slovenia, danilo.dolenc@gov.si

² Regulation (EC) No. 763/2008 of the European Parliament and of the Council on population and housing censuses.

1 Izhodišča, metodološki izzivi, posebnosti in rešitve

1.1 Popolna odvisnost od podatkov, ki so na voljo v virih, ter njihove kakovosti

Prva naloga pri pripravi na prvo izvedbo registrskega popisa v Sloveniji (referenčni datum 1. januar 2011) je bila pripraviti nabor vseh virov, njihovih vsebin ter konceptov, določiti povezovalne identifikatorje ter predvsem oceniti kakovost podatkov posameznih virov, pri čemer smo upoštevali naslednje vidike kakovosti:

- formalna pravilnost podatkov
- ustreznost podatkov za uporabo
- administrativni : statistični koncepti
- popolnost zajetja enot opazovanja.

Ker imamo na SURS dolgoletne izkušnje z uporabo podatkov administrativnih in statističnih virov (predvsem iz Popisa 2002³), smo največ časa in pozornosti namenili dvema novima viroma, ki sta bila v statističnem procesu uporabljena prvič: Evidenca gospodinjstev in Register nepremičnin, ter ključnim podatkom, pomembnim za povezovanje, predvsem številki stanovanja v CRP. Na drugi strani pa smo se soočili z dejstvom, da nekaterih tradicionalnih popisnih vsebin (narodnost, veroizpoved, dnevna migracija na delo oz. v šolo) uporabnikom ne bomo mogli več zagotoviti.

1.1.1 Kakovost podatkov Evidence gospodinjstev

Ker so se podatki o t.i. skupnem gospodinjstvu veliko uporabljali v administrativne namene (sprva so se vodili samo ročno na posebnih kartonih), je bila njihova kakovost tudi po prenosu v e-obliko zelo dobra, lahko rečemo odlična, kljub temu da pri vnosu podatkov upravljavec (Ministrstvo za notranje zadeve – MNZ oz. upravne enote) ni uporabljal nobenih kontrol formalne ali logične pravilnosti. Podatke smo zelo natančno analizirali in ugotovili, da je število formalnih napak zelo majhno (z izjemo ene upravne enote kot posledica izgube papirne dokumentacije). V tesnem sodelovanju z MNZ se je kakovost podatkov do izvedbe popisa 2011 še nekoliko izboljšala, čeprav ta evidenca pri njihovem delu nima prioritete, še zlasti zato, ker bo z letom 2020 ukinjena.

1.1.2 Kakovost podatkov Registra nepremičnin (REN)

Z vidika kakovosti je bila ključna že izvedba popisa nepremičnin v organizaciji Geodetske uprave (GURS), ki je trajal skoraj dve leti, ter seveda dejstvo, da se je register na novo vzpostavil in da se je prvič uporabil prav v popisu prebivalstva. Kot v Revizijskem poročilu ugotavlja Računsko sodišče, je GURS načrtovani časovni okvir za vzpostavitev REN presegel za 63 %. Številne pomanjkljivosti REN so se pokazale šele pri povezovanju teh podatkov s podatki o prebivalstvu. SURS je v letih 2009–2010 intenzivno analiziral podatke REN ter upravljavca vira sproti opozarjal na napake, ki pa jih v tako kratkem času seveda ni bilo mogoče odpraviti. Vse nepravilnosti so se pokazale ob razgrnitvi podatkov oz. obveščanju lastnikov v okviru projekta vrednotenja nepremičnin za potrebe obdavčenja, kar

³ Popis prebivalstva, gospodinjstev in stanovanj v Republiki Sloveniji v letu 2002.

pa je bilo izvedeno šele v letu 2013. Zato pa so bili podatki o stanovanjih v registrskem popisu 2015 že neprimerno boljše kakovosti, kar je ponovni dokaz, da se kakovost administrativnih podatkov izboljšuje samo z njihovo vsestransko uporabo.

1.1.3 Kakovost podatkov o številki stanovanja v CRP

Podatek o številki stanovanja se je zbiral v okviru popisa nepremičnin in bi moral biti prepisan v Centralni register prebivalstva kot sestavni del naslova posameznika v večstanovanjski stavbi. Ob testnem povezovanju podatkov smo ugotovili, da v večstanovanjskih stavbah podatek o številki stanovanja v CRP manjka skoraj za četrtno prebivalcev. Težava je bila tudi to, da ni bilo natančne metodološke opredelitve večstanovanjske stavbe oz. navedbe, v katerih stavbah mora biti oštevilčenje delov stavbe tudi izvedeno. Za izboljšanje zajetja podatka o številki stanovanja so bile sprejete naslednje rešitve, ki jih je izvedel MNZ neposredno v CRP (pri čemer je lahko tudi prišlo do kakšne napake, ki se še vedno odpravljajo):

- avtomatsko določanje številke stanovanja na podlagi lastništva stanovanja in prebivanja na istem naslovu;
- prepis podatka vsem osebam, ki živijo v istem gospodinjstvu z osebo, ki je imela vpisano številko stanovanja;
- skupaj z MNZ izvedeno neposredno zbiranje podatkov od stanovalcev večstanovanjskih stavb (določeno po kriterijih SURS) z uporabo poštne metode - posredovano skoraj 50.000 pošilk, stopnja odgovora je bila 61 %.

Kljub temu je v registrskem popisu 2011 za več kot 12 % oseb v večstanovanjskih stavbah še vedno manjkala številka stanovanja, dodatno pa je bilo 12 % številke stanovanj spremenjenih (popravljenih) zaradi uskladitve z Registrom nepremičnin oz. različnih napak, ki so bile tudi posledica prevelikega časovnega zamika med izvedbo popisa ter vnosom podatkov v CRP ter kasnejših popravkov v REN. Se je pa kakovost podatka o številki stanovanja v CRP izjemno izboljšala do registrskega popisa 2015, saj je skupna stopnja popravkov številke stanovanja v večstanovanjskih stavbah znašala samo še 6,5 %, tako da številke stanovanja ni imelo vpisane samo nekaj več kot 30.000 prebivalcev (v letu 2011 pa trikrat več – skoraj 90.000 oseb).

1.2 Popolna uskladitev popisnih podatkov z redno demografsko statistiko

V zadnjem terenskem popisu (2002) je bila ta razlika 1,6 %. Podatki pripravljene v skladu z novo definicijo prebivalstva, ki je harmonizirana z evropsko Uredbo, so v bistvu osnovni vhodni podatki registrskega popisa. Za potrebe popisa je bilo treba predhodno statistično »urediti« samo osebe, mlajše od 15 let, ki so bile same na naslovu prebivališča (praviloma smo jih statistično preselili k enemu od staršev). Gre za statistično nepomembno število (okrog 300) zapisov.

1.3 Novi koncept gospodinjstva

1.3.1 Koncept in omejitve pri metodološki opredelitvi in statistični izpeljavi

Prebivališče v skladu z novo definicijo prebivalstva, prebivanje v istem stanovanju (ista številka stanovanja v večstanovanjskih stavbah) ter opredeljeno gospodinjstvo iz evidence gospodinjstev (osebe z isto zaporedno številko gospodinjstva na istem naslovu) je novi koncept gospodinjstva v registrskem popisu. S tem smo nadomestili prejšnjo definicijo, ki je temeljila na samoopredelitvi glede na skupno porabljanje sredstev za osnovne življenjske potrebe (stanovanje, hrano, druge nujne življenjske potrebščine ipd.). Pri tem pa je treba poudariti, da opredeljevanje gospodinjstva pri administrativnem organu (upravne enote) poteka zelo podobno, kot se respondenti o gospodinjstvu opredelijo v vzorčnih statističnih raziskovanjih.

Omejitve pri metodološki opredelitvi gospodinjstva:

- podatek o številki gospodinjstva se vodi samo za stalno prebivališče (podatka torej nimajo tujci z začasnim prebivališčem ter osebe, ki imajo istočasno stalno in začasno prebivališče, na katerem jih v skladu z novo definicijo prebivalstva tudi upoštevamo kot prebivalce);
- zakonodaja ne predvideva skupinskih gospodinjstev.

V popisu smo lahko uporabili t.i. koncept gospodinjске enote (housekeeping), ki dopušča, da v posameznem stanovanju prebiva več kot eno gospodinjstvo, kar je v Sloveniji zelo pogosto, predvsem v individualnih hišah.

1.3.2 Metodološke in procesne rešitve za določanje zasebnih gospodinjstev

Za določanje manjkajočih števil gospodinjstva (2,1 % v registrskem popisu 2011) smo pripravili zelo zahtevne avtomatske procedure, ki so temeljile na določenih predpostavkah in pogojih, ki so med drugim zajemali:

- prepis številke gospodinjstva iz stalnega naslova;
- pripis številke gospodinjstva na osnovi znane številke stanovanja;
- določanje številke gospodinjstva na osnovi povezav med starši in otroki ter zakonci preko statističnih identifikatorjev.

Za izboljšanje kakovosti podatkov o gospodinjstvih je bil uporabljen vmesnik za ročne popravke avtomatsko določenih števil gospodinjstev ter določanje števil gospodinjstev za mlajše od 18 let (brez podatka o identifikatorjih staršev) na osnovi priimkov. Ročno je bilo določenih samo 0,7 % števil gospodinjstev. Skupna stopnja popravkov številke zasebnega gospodinjstva je znašala 5,1 %, kar pa je celo nekoliko manj, kot je ta kazalnik znašal v registrskem popisu 2015 (6,5 %). Vendar v tem primeru razlog za to ni poslabšanje kakovosti vhodnih podatkov, temveč večji delež tujcev z začasnim prebivališčem, ki tega podatka v administrativnem viru sploh nimajo zabeleženega.

1.3.3 Metodološke in procesne rešitve za določanje skupinskih gospodinjstev oz. skupinskih stanovanj

Osnovna izhodišča za določanje skupinskih gospodinjstev/stanovanj so bila:

- prebivanje večjega števila oseb v istem stanovanju;
- prebivanje večjega števila oseb z začasnim prebivališčem na naslovu;
- prebivanje večjega števila oseb v gospodinjstvu, ki med seboj niso sorodniki.

Na tak način določene naslove smo najprej uparili s Poslovnim registrom Slovenije oz. jih poiskali preko spleta ter na osnovi opisa določili t.i. tip nastanitve, ki smo ga razdelili v naslednje večje skupine:

- domovi za starostnike
- študentski domovi
- socialno varstveni zavodi
- verske institucije (samostani ipd.)
- samski domovi
- drugi nastanitveni objekti (npr. hoteli).

V naslednji fazi smo v skladu z mednarodnimi priporočili določili, v katerih tipih nastanitve so skupinska gospodinjstva (domovi za starostnike, študentski domovi, socialno varstveni zavodi, samostani) in v katerih zasebna gospodinjstva (samski domovi, drugi nastanitveni objekti). Na vseh teh naslovih pa smo statistično določili tudi skupinsko stanovanje (ne glede na to, da ima katero od teh lahko tudi značilnosti običajnega stanovanja).

Trenutno imamo v bazi okrog 1.500 naslovov posebnih nastanitev, ki jih redno četrtletno ažuriramo, saj se njihovo število tudi zaradi staranja prebivalstva (novi domovi za upokoјence), odpiranja bivalnih enot socialnovarstvenih zavodov, novih nastanitvenih zmogljivosti za študente ipd. povečuje.

1.3.4 Prebivalci z administrativnim naslovom

Zakon o prijavi prebivališča omogoča nekaterim posebnim skupinam prebivalstva (npr. t.i. brezdomcem) prijavo prebivanja na naslovih institucij kot so centri za socialno delo ali karitativne organizacije. Seveda te osebe na teh naslovih dejansko ne prebivajo, praviloma na teh naslovih tudi ni običajnih stanovanj. Podobno kot za skupinska gospodinjstva smo za te osebe določili nov tip gospodinjstva, ki smo ga poimenovali posebno gospodinjstvo. V popisu 2011 je bilo takih oseb 1.400, v letu 2015 pa že skoraj 3.000. Tudi ta gospodinjstva so bila določena na podoben način kot prava skupinska gospodinjstva, največja razlika pa je, da je na nekaterih naslovih zelo majhno število administrativno prijavljenih oseb.

1.3.5 Implementacija v vzorčna raziskovanja oseb/gospodinjstev

Določitev skupinskih ter posebnih gospodinjstev oz. naslovov, na katerih se ta nahajajo, je omogočila, da smo lahko začeli v vzorčni okvir vključevati le osebe, ki živijo v zasebnih gospodinjstvih, kar je tudi osnovni kriterij pri zbiranju vzorčnih raziskovanj, ki se izvajajo v skladu z evropskimi uredbami. Predvsem pa smo s tem že pred začetkom terenskega dela izločili neustrezne enote opazovanja in s tem zmanjšali stroške ter racionalizirali terensko izvedbo raziskovanj v organizacijskem smislu.

1.4 Generiranje podatkov o družinah

1.4.1 Koncept in omejitve pri metodološki opredelitvi in statistični izpeljavi

Podatki o družinah v gospodinjstvu so metodološko izpeljani na isti način kot v preteklih terenskih popisih. Bistvena razlika je v tem, da je zdaj večina podatkov o družinah izpeljana avtomatsko na osnovi:

- podatkov o razmerju do referenčne osebe gospodinjstva;
- identifikatorjev staršev oz. zakoncev.

V terenskih popisih so se podatki o družinah do leta 1991 kodirali ročno v fazi priprave popisnega gradiva, v Popisu 2002 pa so jih kodirali popisovalci na terenu. Oba načina sta se z vidika kakovosti generiranja podatkov o družinah izkazala za problematična, tudi zaradi velikega števila vključenih v ta proces in posledično neustrezne usposobljenosti za to relativno zahtevno metodološko opravilo.

Najpomembnejše omejitve pri avtomatski izpeljavi podatkov o družinah so bile:

- identifikatorji staršev/zakoncev niso na voljo za starejše generacije ter večino tujcev z začasnimi prebivališčem;
- nekatera razmerja niso dovolj povedna (drug sorodnik, nesorodnik – na ta način je v evidenci gospodinjstev pogosto opredeljen zunajzakonski partner otroka referenčne osebe gospodinjstva);
- posamezne osebe v gospodinjstvu lahko teoretično sodijo v več družin;
- razmerja v gospodinjstvih z večjim številom članov so zapletena.

1.4.2 Metodološke in procesne rešitve za generiranje družin

V prvi fazi avtomatske izpeljave podatkov o družinah so bile uporabljene naslednje rešitve, s katerimi so se najprej uredila razmerja do referenčne osebe gospodinjstva:

- urejanje gospodinjstev z več kot eno referenčno osebo;
- urejanje gospodinjstev brez referenčne osebe;
- preverjanje pravilnosti razmerja do referenčne osebe, če so na voljo identifikatorji staršev oz. zakonca.

V naslednji fazi so se generirali podatki o družinah po naslednjem zaporedju:

- za vsako kombinacijo razmerij do referenčne osebe gospodinjstva, ki omogoča enolično izpeljavo podatkov o družinah v gospodinjstvu, so bile pripravljene matrike (skoraj 200) za izpeljavo položaja v družini ter tipa družine;
- za izpeljavo bolj kompleksnih razmerij (npr. vnuk v gospodinjstvu bi lahko pripadal dvema sestrama – otrokoma referenčne osebe) so bili poleg matrik upoštevani še podatki o identifikatorju staršev;
- avtomatsko so bile statistično določene zunajzakonske skupnosti v dvočlanskih gospodinjstvih;
- ročno urejanje podatkov o družinah preko vmesnika, kjer se podatki niso v celoti ali sploh ne avtomatsko izpeljali;
- ročno popraviljanje avtomatsko izpeljanih podatkov o družinah z logičnimi napakami (predvsem preverjanje pravilnosti razlik v starosti med člani družine).

Ročno je bilo določenih ali popravljenih 0,9 % podatkov o razmerju do referenčne osebe gospodinjstva. Skupaj je stopnja imputiranja ter popravkov teh podatkov znašala 8 %. Tudi v registrskem popisu 2015 je bila skupna stopnja popravka podatka o razmerju do referenčne osebe gospodinjstva skoraj povsem enaka (7,5 %), kar je logično, saj se tako kakovost vhodnih podatkov kot metodologija izpeljave teh podatkov nista spreminjali.

1.5 Metodologija izpeljave popisnih vsebin iz večjega števila virov

Medtem ko smo osnovne demografske vsebine, podatke o gospodinjstvih in družinah ter podatke o stanovanjih prevzeli iz samo enega vira, pa je bilo za nekatere popisne vsebine uporabljenih večje število virov, največ za podatke o izobrazbi (11), ker v Sloveniji ni na voljo centraliziranega registra izobrazbe (čeprav je ta spremenljivka navedena kot vsebina v Zakonu o CRP), ter statusu aktivnosti (8), kar pa je v prvi vrsti statistični koncept. Po štiri različne vire podatkov smo uporabili tudi pri pripravi podatkov o selitvah ter številu živorojenih otrok.

1.5.1 Izobrazba in status aktivnosti

Podatki o izobrazbi in aktivnosti celotnega prebivalstva so bili pripravljene po metodologiji hierarhije vira. Vsak predviden vir podatkov smo predhodno pregledali ter ocenili njegovo primernost za uporabo z vidika kakovosti oz. zanesljivosti, ažurnosti, dostopnosti ter primerljivosti podatkov. Kadar je podatek o statusu aktivnosti za isto osebo na voljo (dostopen) v več virih, se po metodologiji hierarhije vira upošteva tisti podatek, ki je v viru, označenem z višjo stopnjo prioritete (ne glede na to, ali je podatek v več virih enak ali različen). Pri določanju prioritete virov za status aktivnosti smo izhajali tudi iz hierarhije statusov aktivnosti ter mednarodnih priporočil za to področje. Prednost imajo viri s podatki o delovni aktivnosti, sledijo jim viri s podatki o osebah, ki so vključene v izobraževanje, nazadnje pa se uporabijo viri s podatki o neaktivnem prebivalstvu.

Priprava podatkov o najvišji dokončani izobrazbi je še poseben zahteven statistični proces, saj se zbirajo podatki o osebah, ki so se izobraževale v različnih obdobjih in v različnih izobraževalnih sistemih. Poleg tega je treba upoštevati, da se vedno večji del prebivalstva med zbiranjem podatkov še izobražuje. Zato je bilo predhodno treba vse podatkovne vire o izobrazbi (baza o diplomantih terciarnega izobraževanja tako sega celo v leto 1989) prevesti na skupno klasifikacijo (KLASIUS–SRV). Če je bilo v posameznem viru za isto osebo na voljo več podatkov o izobrazbi, smo upoštevali najvišjo raven dosežene izobrazbe (če je oseba po visokošolski univerzitetni izobrazbi dokončala še magisterij znanosti, smo upoštevali magisterij znanosti).

Metodološke rešitve smo morali najti tudi za primere, ko smo za posamezno osebo imeli na voljo vsaj dva podatka o izobrazbi, pri čemer je bila najvišja dosežena izobrazba v viru z nižjo prioriteto (in s tem v viru, ki je manj kredibilen) višja od podatka, ki smo ga našli v bolj kakovostnem viru.

Poleg tega smo nekatere podatke prevzeli še iz Popisa 2002, in sicer podatke o izobrazbi oseb, ki jih nismo našli v nobenem od predhodnih desetih virov glede na prioriteto (praviloma

so to podatki o starejših osebah, ki se po letu 2002 niso več izobraževale ali niso bile delovno aktivne po letu 1990).

1.5.2 Meddržavne in notranje selitve

Posebnost dela teh podatkov je, da so nespremenljivi (država oz. kraj rojstva), določeni dogodki pa so lahko pogosti oz. posameznika doletijo večkrat v življenju (predvsem notranje selitve). Z vidika daljše časovne perspektive pa so v tem kontekstu zanimivi tudi podatki o prebivanju v tujini (kar je tudi del vsebine evropske Uredbe). Zato je zelo pomembno, da so za podatke o selitveni zgodovini posameznika na voljo čim daljše serije podatkov oz. da se lahko določene podatke prevzame iz predhodnega popisa ter se jih samo ažurira s spremembami med dvema izvedbama. Ker gre predvsem za teritorialne podatke (z izjemo podatkov o letu dogodka), je zaradi sprememb upravno-teritorialne ureditve (npr. nastanek novih občin) oz. nastanka novih držav (za nas je najbolj pomemben razpad nekdanje Jugoslavije) treba podatke predhodno ustrezno prekodirati na obstoječe stanje teritorija ob popisu.

1.5.3 Rodnost (živorojeni otroci)

Posebnost izpeljave teh podatkov je, da imamo opravka z generacijami, ki v veliki meri (še) niso rodile, tistimi v rodni dobi ter z generacijami, ki so rodno dobo že zaključile, kar pomeni, da se podatki za te ženske ne bodo več spreminjali. Ker uporabljeni viri predvsem za starejše generacije ter priseljene ženske (ki so lahko rodile že v izvornem okolju) niso popolni, smo načeloma uporabili metodologijo najvišje vrednosti podatka v uporabljenih virih. Če je bila vrednost podatka v več virih enaka, eden izmed teh virov pa je bil CRP, smo kot vir upoštevali CRP, ki se je tudi izkazal kot najbolj relevanten vir za te podatke. Iz CRP smo želene podatke namreč izpeljali na osnovi bioloških vezi med materjo in otrokom (otroci). Iz rednega statističnega raziskovanja o rojenih smo prevzeli podatek o številu vseh živorojenih otrok – če je ženska med letoma 2002 in 2010 rodila večkrat, smo upoštevali podatek o številu otrok pri navedbi zadnjega poroda. Iz Popisa 2002 smo prevzeli ta podatek za tiste ženske, za katere tega podatka v CRP ni bilo ali pa je bila vrednost tega podatka v popisu 2002 višja. Iz Popisa 2002 so bili ti podatki prevzeti pretežno za najstarejše prebivalke Slovenije. Za ženske, ki niso rodile v Sloveniji, tu pa živijo s svojimi otroki, smo manjkajoči podatek pridobili tako, da smo upoštevali število otrok, ki živijo z njimi v družini.

1.6 Metodološke novosti registrskega popisa

1.6.1 Generacijska sestava gospodinjstev

Generacije v gospodinjstvih so bile izpeljane avtomatsko, s pomočjo matrik za izpeljavo podatkov o družinah. Za vsako kombinacijo razmerja do referenčne osebe gospodinjstva je bila določena tudi generacija posameznika v gospodinjstvu. Pri tem gre za relativno enostavno izpeljavo teh podatkov, ker je gospodinjstev z bolj zapleteno sestavo (tri ali štiri generacije, gospodinjstva z nezaporednimi generacijami) relativno zelo malo (7,5 %). Štiri zaporedne generacije so živele v nekaj več kot 1.000 gospodinjstvih.

1.6.2 Povezave med starši in otroki

Identifikatorji staršev so statistično zelo dober podatek predvsem za mlajše generacije. Tako je imelo podatek o materi 98,3 %, podatek o očetu pa 97 % otrok, mlajših od 18 let. Do starosti 50 let sta ta dva deleža bila še vedno dovolj visoka za relevantnost teh podatkov in nadaljnje izpeljave podatkov na tej osnovi (mati: 91 %, oče: 88 %). Podatki o starših so torej na voljo za generacije, ki so za raziskovalno in analitsko delo najbolj zanimive. Predvsem raziskovalcem socialnih razmerij ter priseljenkega ozadja (angl. immigrant background) so namenjeni podatki, ki so izpeljani iz identifikatorjev otrok ter njihovih staršev:

- status matere in očeta (živita v isti družini, istem gospodinjstvu, drugje v Sloveniji, nista prebivalca Slovenije);
- vzpostavljene družine (vsaj en otrok v družini ni skupen biološki otrok obeh staršev);
- socio-ekonomske in selitvene značilnosti staršev (izobrazba, status aktivnosti, prvo prebivališče oz. država izvora);
- status priseljenosti (generacije priseljencev ter tipi priseljencev glede na območje izvora).

2 Evalvacija kakovosti registrskih podatkov o prebivalcih Slovenije, gospodinjstvih in družinah

V tem prispevku bomo obravnavali tri vidike kakovosti registrskih podatkov:

- notranja distribucija prebivalstva kot posledica razlik med dejanskim in registriranim prebivanjem;
- precenjenost oz. podcenjenost registra prebivalstva in vpliv tega na statistične, demografske podatke;
- vpliv registrskih podatkov na sestavo gospodinjstev/družin.

2.1 Metode za evalvacijo kakovosti registrskih podatkov o prebivalcih

Precenjenost registra je splošen problem vseh registrskih držav, saj so predvsem podatki o odseljevanju v tujino praviloma podcenjeni oz. se registracija teh dogodkov dogaja z določenim časovnim zamikom. Precenjenost podatkov registra in tudi razlike med dejanskim in registriranim prebivanjem lahko ugotovljamo v terenskih raziskovanjih oseb in gospodinjstev, v katerih zbiramo podatke o tem, kakšen je status v vzorec izbrane osebe:

- živi drugje v Sloveniji (praviloma na drugem naslovu nima registriranega prebivališča),
- živi v tujini (precenjenost registra),
- je umrla (posledica zamika med pripravo vzorca in anketiranjem),
- neznano (anketar iz različnih vzrokov ni mogel vzpostaviti stika oz. ugotoviti statusa izbrane osebe).

Podatki iz terenskih raziskovanj APG (poraba gospodinjstev), SILC (življenjski pogoji), ANP (delovna sila) in IKT (uporaba interneta) kažejo, da drugje v Sloveniji (praviloma na naslovu,

na katerem nimajo prijavljenega prebivališča) živi 4–5 % prebivalcev, da jih od 1–2 % živi v tujini, da za manj kot 1 % oseb statusa ni bilo mogoče določiti, medtem ko je delež umrlih povsem skladen s stopnjami umrljivosti oz. zanemarljiv. Žal se temu, da v vzorec izberemo tudi umrle osebe, ne moremo izogniti zaradi časovnega zamika med pripravo vzorca in anketiranjem.

Ko pa iz podatkov raziskovanj o gospodinjstvih (APG, SILC, ANP) izpeljemo ravno nasproten pristop – koliko oseb, ki nimajo registriranega prebivališča na naslovu izbrane osebe, živijo pa v anketiranem gospodinjstvu –, pričakovano dobimo nekoliko nižji delež (do 3 %), kar je posredni indikator notranje distribucije prebivalstva.

Posebna metoda za ugotavljanje precenjenosti registra pa je delež imputiranih vrednosti za status aktivnosti v registrskem popisu. To pomeni, da je podatek o posamezniku na voljo samo iz CRP (kjer ima oseba veljavno prebivališče v Sloveniji ter je v skladu s statistično definicijo običajnega prebivališča opredeljena kot prebivalec Slovenije), ni pa podatka o tej osebi v nobenem od osmih virov za določanje statusa aktivnosti. Ker podatek o statusu aktivnosti po metodologiji registrskega popisa izpeljujemo vsako leto na stanje 1.1., ugotavljamo, da se delež imputiranih vrednosti zmanjšuje (1,5 % v letu 2011; 1,3 % v letu 2015). Iz leta v leto se zmanjšuje tudi število oseb, ki jim imputirane vrednosti pripišemo iz predhodnega leta, kar kaže na izboljšanje kakovosti podatkov o dejanskem prebivanju v Sloveniji oz. urejanje statusa prebivališča v administrativnem viru.

Z vidika precenjenosti podatkov registra so najbolj problematična skupina prebivalcev tujci s stalnim prebivališčem v Sloveniji (1. 1. 2015 je bila stopnja imputiranja statusa aktivnosti za njih 5,3 %). Najvišja stopnja imputiranja sicer velja za tujce z začasnim prebivališčem (9,5 %), ki pa jih statistično odselimo takoj, ko jim izteče dovoljenje za prebivanje v Sloveniji (če seveda prebivanja v Sloveniji ne podaljšajo).

Zmanjšuje pa se število t.i. administrativno preživelih oseb (praviloma so to starejši), ki pa so dejansko mrtve. Če je po oceni za leto 2007 bilo število stoletnikov v Sloveniji precenjeno za 15–20 %, je ta delež v 2015 znašal največ 8,5 % (za 20 od 236 stoletnikov namreč ni bilo podatka o statusu aktivnosti v nobenem od virov). Med starimi 95 ali več let pa je stopnja imputiranja in s tem precenjenosti znašala 3,0 %.

2.2 Izvedba posebnega terenskega raziskovanja o kakovosti podatkov o prebivalstvu Slovenije

Vzorec za terensko raziskovanje je bil pripravljen na osnovi naslednjih hierarhičnih kriterijev (posameznik bi bil lahko izbran po več kriterijih), pri katerih smo upoštevali prebivalce Slovenije na dan 1. 1. 2016 in vir podatka o statusu aktivnosti (imputacija):

- tri zaporedna leta ni podatka o statusu aktivnosti oseb (2013–2015);
- za tuje državljane s stalnim prebivališčem ni podatka o statusu aktivnosti v zadnjem letu (2015);
- za državljane Slovenije s samo začasnim prebivališčem ni podatka o statusu aktivnosti v zadnjem letu (2015).

Poleg tega smo v vzorec za terensko raziskovanje vključili še osebe, ki glede na statistične kriterije ne sodijo med prebivalce Slovenije, saj imajo poleg prijavljenega stalnega prebivališča v Sloveniji tudi začasno prebivališče v tujini (t.i. zdomci), vendar imajo zapisan podatek o statusu aktivnosti, ki kaže na verjetnost, da živijo v Sloveniji.

Kriterijem je ustrezalo nekaj več kot 18.000 oseb, končni vzorec pa je obsegal 14,219 zapisov, saj smo podatkovno bazo pred začetkom raziskovanja (sredina marca 2016) dodatno prečistili (vprašalnik je bil poslan samo eni osebi iz gospodinjstva/stanovanja) ter jo ažurirali z najnovejšimi podatki CRP (stanje 1. 3. 2016). V prvi fazi smo kratek vprašalnik (2 strani) posredovali naslovnikom po pošti, s priloženo kuverto za odgovor in plačano poštnino. Nekaj več kot 11 % pošiljk se je vrnilo neposredno preko izvajalca (Pošta Slovenije) s pripisoma 'Neznan naslovník' oz. 'Preseljen'. Stopnja odgovora je bila po pričakovanju relativno nizka (17,5 %), kar pomeni, da 71 % prejemnikov na naše pismo ni odgovorilo. V drugi fazi pa smo določili nekatera strnjena območja, v katerih smo po neodgovoru na poštno metodo izvedli še terensko raziskovanje (7 anketarjev, metoda PAPI, vzorec 1.915 oseb).

Odgovore respondentov smo razdelili v pet skupin z vidika precenjenosti / podcenjenosti statističnih podatkov o prebivalstvu:

- prebivalec živi v tujini – precenjenost;
- administrativno preživeli prebivalci – precenjenost;
- zdomci – podcenjenost;
- prebivalec živi v Sloveniji – pravilno stanje;
- neznano – ni podatka o tem, kje respondent prebiva.

Če rezultate raziskovanja generaliziramo, potem lahko ugotovimo, da nekaj več kot polovica vzorca sodi med precenjene podatke registra in s tem prebivalstva (dejansko živijo v tujini), manj kot petina med podcenjenost števila prebivalstva, saj dejansko živijo v Sloveniji, ne sodijo pa med prebivalce Slovenije, četrtnina pa je pravilno razvrščenih med prebivalce Slovenije.

Na agregatni ravni je torej število prebivalcev Slovenije precenjeno za približno 0,5 % oz. več kot 10.000 oseb, na mikro ravni pa nekoliko več (0,7–0,8 %). Je pa očitno to raziskovanje izpodbudilo kar precejšnje število posameznikov, ki so si po prejemu vprašalnika uredili svoje prebivališče v tujini oz. odjavili stalno prebivališče v Sloveniji. Zanimivo je, da kar 83 % tistih, ki so to naredili, sploh ni odgovorilo na vprašalnik.

2.3 Nekaj kazalnikov kakovosti podatkov o gospodinjstvih/družinah

Pravilnost registracije dejanskega prebivanja ima največji vpliv na tipe oz. sestavo družin, medtem ko je vpliv na število oz. povprečno velikost gospodinjstev zanemarljiv. Podatki popisov 2011 in 2015 namreč kažejo, da je najbolj precenjeno število enostarševskih družin. Na to nas navajajo podatki, da je četrtnina mater z otroki v enostarševski družini poročenih, prav tako pa tudi podatki, da se večina otrok rodi v zunajzakonski skupnosti. Verjetno tu lahko govorimo tudi o zamiku registracije skupnega prebivanja, ki se pogosto realizira šele z rojstvom otroka oz. potem, ko si uspejo rešiti stanovanjski problem. Na podcenjenost podatkov o skupnem sobivanju nas posredno navajajo tudi podatki o številu naseljenih stanovanj, ki jih je bilo v 2015 skoraj odstotek več kot v popisu 2011, čeprav se je število

stanovanj v istem obdobju povečalo za 9 %. Stanodajalci zaradi oddaje na črno pogosto ne dovolijo prijave začasnega prebivališča v njihovi nepremičnini, zato je posledično tudi število naseljenih stanovanj podcenjeno. Razlogov, da 25.000 parov zakoncev (skupaj več kot 51.000 oseb oz. 7 % poročenih) ni prijavljenih na istem naslovu, pa ne poznamo.

3 Zaključne misli

Prva izvedba registrskega popisa v 2011 v Sloveniji je bila nedvomno nov pomemben razvojni dosežek slovenske statistike. Še večji pomen temu daje dejstvo, da je bil celoten projekt izveden v okviru rednih delovnih obveznosti samo z lastnimi človeškimi viri SURS, brez dodatnega zaposlovanja ali sodelovanja zunanjih sodelavcev. Z drugo izvedbo v letu 2015 smo uporabnikom ponudili tudi metodološko primerljive podatke. Nekatere tradicionalne popisne vsebine (izobrazba, status aktivnosti) so zdaj za celotno prebivalstvo na voljo celo vsako leto.

Vzporedno z razvojem metodologije in izboljšavo (predvsem racionalizacijo) statističnih procesov se povečuje tudi kakovost vhodnih (predvsem administrativnih) virov podatkov. Pri tem lahko le obžalujemo dejstvo, da bomo po letu 2021 izgubili dragocen vir podatkov o gospodinjstvu, saj to evidenco ukinja letos julija sprejeti Zakon o prijavi prebivališča. Bo pa to po drugi strani nov metodološki izziv, ki ga bomo znali brez dvoma uspešno rešiti.

Podatke, ki jih na SURS pripravljamo v okviru socialnih statistik iz administrativnih in drugih virov podatkov, v vse večji meri vključujemo tudi v vzorčna raziskovanja, s čimer prihranimo tako čas za zbiranje in procesiranje (ter seveda tudi stroške) in ne obremenjujemo dajalcev podatkov. Še bolj pomembno pa je, da se s tem povečuje harmonizacija podatkov med posameznimi raziskovanji. Naše izkušnje z zadnjim terenskim popisom v 2002 pa tudi kažejo, da so podatki registrskega popisa bolj kakovostni. Pri tem se seveda zavedamo določenih omejitev teh podatkov ter tudi dejstva, da so podatki nekoliko precenjeni. Terensko raziskovanje, ki smo ga izvedli, pa bo tudi dobra osnova za poglobljeno analizo ter evalvacijo kakovosti podatkov o prebivalstvu ter morebitne metodološke oz. procesne spremembe v prihodnje. Zelo pomembno je tudi, da se upravljavci administrativnih virov zavedajo svoje odgovornosti in zavezanosti kakovosti, ki se odraža predvsem pri nadaljnji uporabi podatkov. Statistika si v skladu s svojo metodologijo lahko dovoli, da registrske podatke spremeni, popravi, dopolni, saj podatke objavlja v agregirani obliki. In prav to je tudi bistvo registrskega popisa – iz surovih, nepovezanih podatkov ponuditi uporabniku užitno in kakovostno statistiko, ki mora v čim večji meri odražati realno stanje.