RAHVALOENDAJATE TEGEVUS KÜSITLUSE JÄREL[a]

Ene-Margit Tiit
Statistikaamet

Mare Vähi
Tartu Ülikool

Rahva ja eluruumide loenduse küsitlus lõppes 31. märtsil. Sellele järgnes mitu kuud pingsat andmekorrastust ning 31. mail saime teada, et Eestis loendati ühtekokku 1 294 236 püsielanikku. Kuidas selle tulemuseni jõuti? Mida tegid ja teevad rahvaloendajad küsitluse järel?

Loendusjärgne andmekorrastus

Loendusjärgne andmekorrastus on ka varasematel loendustel kestnud kuid, mõnikord isegi üle aasta. Mida rohkem on küsimusi ja mida rangemad kvaliteedinõuded, seda rohkem kulub üldiselt aega. Loomulikult kasutatakse tänapäeval andmete korrastamisel arvutite abi, kuid ka siin on piirid – inimeste leidlikkus eksimisel on märksa suurem kui programmeerijad oskavad seda ette näha.

Ajalooliste rahvaloenduste puhul oli ette nähtud ankeetide täitmise ja parandamise kord – loendajad täitsid ankeedid pliiatsiga, vajalikud parandused tehti maha kriipsutades, mitte kustutuskummi abil; loendusjuht tegi parandused musta tindiga ja viimase taseme kontroll – punase tindiga. Parandati peamiselt loogilisi vigu, mis ankeetide läbivaatamisel silma paistsid: vanemate ja laste vanuseerinevused, töö vastavus haridusele ning seisusele jne.

Tänapäeval, kui andmeid kogutakse arvuti ja interneti abil, teeb niisuguse kontrolli programm juba andmete kogumise ajal, juhtides tähelepanu vähetõenäolistele olukordadele (Kas tõesti on 22-aastasel vastajal doktorikraad? Kas tõepoolest on elukaaslaste vanusevahe 35 aastat?). Kui vastusevariant on täiesti võimatu (isa on noorem kui tema poeg), siis ei lase programm ankeedi täitmist jätkata, enne kui viga on parandatud. Esmane andmekorrastus koosneb tänapäeval peamiselt kolmest osast: isikute identifitseerimine, aadresside identifitseerimine ja kodeerimine. Järgneb eluruumide ja leibkondade identifitseerimine ning isikute sidumine nendega.

Isikuankeetide identifitseerimine

Isikuankeetide (isikute) identifitseerimine kõlab kahtlaselt: kas see tõepoolest tähendab võimalust tutvuda isikustatud isikuandmetega? Nii see siiski ei ole. Kui kõik vastajad ja küsitlejad töötaksid ideaalselt hästi, siis poleks seda etappi üldse tarvis. Kahjuks aga laekus suur osa isikuankeetidest ilma isikukoodita – internetiloendusel hinnanguliselt kümnendik ja küsitlusel koguni üle kolmandiku. Korraldajatele oli ootamatu, et nimelt küsitlusel jättis suur hulk loendatavaid oma isikukoodi ütlemata. Oli isegi kuulda arvamusi, et loendatavad ei julgenud oma isikukoodi öelda, kartes, et keegi kasutab seda SMS-laenu võtmiseks!

Isikukoode on aga kindlasti tarvis kahel otstarbel. Esiteks seotakse isikukoodi abil loendusankeedi andmed andmetega, mida loendusel ei küsitud, vaid võeti mõnest registrist (nt õppimine), samuti andmelünkade täitmiseks registriandmetega. Teiseks on isikukoodid tarvilikud selleks, et andmestikust kõrvaldada duplikaatankeedid (kui ühe isiku kohta on laekunud mitu ankeeti).

Üks aegaviitvamaid tegevusi esmase andmekorrastuse protsessi ajal ongi isikukoodide lisamine ankeetidele. Nii suures osas, kui võimalik, toimub see automaatselt rahvastikuregistri abil, kusjuures aluseks on isiku nimi ja sünniaeg. Kui need andmed on õigesti märgitud, on ülesanne

[a] Artikli esialgne lühem variant ilmus 11.05.2012 ajalehes Sirp – E.-M. Tiit „Pärast loendust loenduse pärast“.

võrdlemisi lihtne: keskmiselt on iga päev Eestis sündinud ca 50 inimest, nende hulka juhtub sama ees- ja perekonnanimega inimesi võrdlemisi harva (eriti eestlaste puhul, sest meil on kasutusel olevate ees- ja perekonnanimede loetelu üsna pikk). Kui see juhtub, kasutatakse lisateavet, näiteks elukoha andmeid.

Asi läheb keeruliseks aga siis, kui lähteandmed pole õiged. Sünniaja puhul võib olla eksitus nii päevas kui ka aastas. Nimede puhul on eksimisvõimalusi palju. Kuulmise järgi üles kirjutades võivad segi minna Mati ja Matti, Avo ja Aavo, Kristiina ja Kristina, Anneli ja Annely, seda enam, et iga kord ei tarvitse loendajale vastav leibkonnaliigegi isiku täpset nimekuju teada, kui igapäevakasutuses on hoopis hüüdnimi. Veelgi enam variante tuleb slaavi või mõnes muus ladina tähestikust erinevas tähestikus originaalnime transkribeerimisel Eesti rahvastikuregistris esinevale nimekujule, mis on isikukoodi leidmiseks vajalik. Näiteks võib tuua perenime kujud Kholodny, Kholodniy, Holodnõi. Nii tekivadki samastamisülesanded, mida programm alati automaatselt lahendada ei suuda ja kus appi peavad tulema operaatorid, kes näevad isiku mõningaid täiendavaid andmeid ankeedist ja samu andmeid rahvastikuregistrist, et neid siis vajaduse korral kokku sobitada. See töö lõppes põhilises osas kuu aja jooksul pärast loendust.

Kellelgi pole siiski tarvis muretseda selle pärast, et identifitseerimise käigus saavad operaatorid ankeete isikustatud kujul vaadata. Operaator näeb ankeedist ainult oma tööks hädavajalikku osa, hiljem aga analüüsitakse ankeete hoopiski ilma nimeta ja ka isikukoodid on teisendatud (krüptitud) nii, et neid konkreetse isikuga siduda pole võimalik. Küll aga võimaldavad krüptitud koodid isikute andmeid siduda registrite andmetega, mis on samuti kasutatavad anonüümseks muudetud kujul.

Aadresside identifitseerimine

Olukord pole lihtsam ka aadressidega. Eestis on olemas aadressi standardkuju, mis peaks võimaldama kõiki eluruume tuvastada. See kuju oli kasutusel ka loendusankeetides. Kuid loendatavatel oli võimalik standardkujust mööda minna ja kasutada lisavariandina n.ö vaba käega kirjutatud aadresse. Nii ongi standardse aadressi “Harjumaa, Tallinna linn, Kesklinna rajoon, Endla tänav, maja nr 15“ kõrval ka aadress „Tallinn, Endla 15“ ja veel mitmeid muid aadressikujusid. Selleks, et iga eluruum oleks andmestikus nõuetekohaselt esindatud ja näiteks nurgamajad ei esineks andmestikus kahe hoonena, on tarvis ka aadressiandmed korrastada. Taas on oluline osa sellest tööst tehtav programmiliselt, kuid keerukamad probleemid peavad lahendama operaatorid. Aadresside osas on selle loenduse puhul tööd ja probleeme rohkem kui varasematel loendustel, sest iseloendamisel (internetis) kirjutasid inimesed oma aadressi üles ise ja seetõttu on varieerumisvõimalusi märksa rohkem, kui oli varasematel loendustel, mil aadressi kirjutasid loendajad, keda selleks oli vajalikul viisil juhendatud. Lisaprobleeme tekitab see, et vahepeal kadusid käibelt talunimed (hoone nimed), mis praegu kuuluvad standardse aadressi koosseisu kohtadel, kus ei ole nimedega varustatud tänavad. Ülesande muudab keerukamaks (aga tulemuse loodetavasti märksa täpsemaks) ka elukoha märkimine kaardil, mida sellisel kujul tehti Eestis esimest korda ja mis oli iseloendusena võrdlemisi uudne kogu maailmas. Uuenduslik oli ka see, et niihästi iga leibkond kui ka iga inimene võis märkida enesele kaks elukohta – peamise ja teisese. Seegi suurendab loendatavate eluruumide arvu ja vastavalt ka tööd aadresside täpsustamisel.

Kodeerimine

Kolmas töö, mida operaatorid pärast loendust on teinud, on kodeerimine. Lihtsustatult tähendab see tekstiliste ankeedivastuste asendamist numbrite ehk nn koodidega. See puudutab eeskätt ameteid ja tegevusvaldkondi, samuti osaliselt murrakuid, uskusid jmt. Rahvusvaheliselt on välja töötatud klassifikaatorid niihästi ametite kui ka tegevusvaldkondade jaoks, kusjuures need klassifikaatorid, mis sisaldavad tuhandeid detailselt kirjeldatud alajaotusi, on universaalsed kogu maailmas, hõlmates nii kartulite kui ka banaanide kasvatajaid, kooli- ja lasteaiaõpetajaid, kirikuõpetajaid ja vangivalvureid, ministreid ja kojamehi.

Need klassifikaatorid on aja jooksul märgatavalt muutunud vastavalt sellele, kuidas uusi ameteid ja tegevusvaldkondi on juurde tulnud. On ka ameteid, mida tänapäeval enam ei tunta või mis on uue vormi saanud – näiteks voorimees, mis esimese rahvaloenduse ajal oli levinud amet Tallinna eesti soost meeste seas. Kõigi loenduste puhul on elukutsete ja ametite tüüpliigendus olemas olnud ja mõlemaid on alati sobivatesse klassidesse sorteeritud. Üks huvitavamaid järeldusi, mida varasemate rahvaloenduste andmete põhjal on võimalik teha, ongi erinevate ametite sagedus ühiskonnas ja selle muutumine. Esimeste Eestis toimunud rahvaloenduste ajal olid n.ö valgekraelised ametid valdavalt sakslaste kanda (arstid, õpetajad, ametnikud, aga ka näiteks kellassepad), militaarseid täitsid venelased ja eestlaste osaks oli peale põllutöö ka töö (juba loodud) vabrikutes, suur osa käsitöödest (kingsepad, pagarid, õmblejad), aga ka teenijaks olemine. Viie esimese rahvaloenduse andmetel oli eestlaste põhiliseks tegevusvaldkonnaks põllumajandus ja kalapüük.

Duplikaatide töötlus

Järgmine samm, mis loendusandmetega tehakse, on duplikaatide töötlus. Üks osa isikuid on loendatud korduvalt: näiteks on isiku ankeedi täitnud abikaasa, laps ja lõpuks ka ta ise. Ise mitu korda ankeeti täita polnud võimalik, kuid isikuid, kes teiste pereliikmete ankeete täitsid, võis olla mitu. Oli ka võimalik, et mingil põhjusel täideti isikule üks ankeet e-loenduse, teine küsitluse ajal. Nii tekkisid duplikaatankeedid, mille hulgast tuli välja valida kõige õigem, nn originaal, ja ülejäänud arvestusest välja jätta. Kuigi meie hea kolleeg arvutiteaduse instituudist arvas, et isikukoodide olemasolul on unikaalsete ankeetide hulga määramine ühe hiireklikiga teostatav, siis tegelikult on asi siiski märksa keerukam.

Lõppenud loendusel koguti isikuandmeid mitmesuguste ankeetide ja dokumentide abil. Püsielaniku ankeetide kõrval (püsielanikud moodustavadki Eesti rahvaarvu) olid kasutusel veel ajutiste elanike ankeedid, mida täideti niihästi isikute kohta, kelle püsielukoht on Eestis, kui ka nende kohta, kes elavad püsivalt välismaal, kusjuures ankeete täitsid enamasti leibkonnaliikmed, kuid võimalik oli ka ise täita. Eestist lahkunute ankeete paluti täita lähisugulastel. Mõnest asutusest (nt sügava puudega isikute haigla või hooldekodu) laekusid loendamisele kuuluvate isikute nimekirjad. Olemas on ka nende isikute isikukoodide loetelu, kes püüdsid end välismaal loendada, kuid kelle loendamine peatati, sest nad märkisid oma püsielukoha riigiks välismaa. Mõnesid isikuid, kes olid loendamata jäänud, kuid enesest vahetult pärast loenduse lõppu teada andsid, intervjueeriti telefonitsi. Iga sellise isikute kategooria kohta on selge, kas nad loetakse Eesti püsielanike hulka või mitte. Nii ei ole Eesti püsielanikud need, kes on siit (omaste kinnitusel) lahkunud, samuti välismaal püsivalt elavad isikud (hoolimata sellest, et nad ennast soovisid loendada) ja ka need isikud, kes elavad Eestis ajutiselt (3–12 kuud), kuid kelle püsielukoht on välismaal.

Asja teeb aga keeruliseks see, et enam kui 100 000 isiku kohta on duplikaatankeete või dokumente ja need on pahatihti vasturääkivad. Näiteks on inimene püüdnud ennast loendada välismaal, loendamine on katkestatud, kuid seejärel on inimene tulnud Eestisse ja täitnud siin oma käega püsielaniku ankeedi. Üsna tüüpiline on ka olukord, kus isiku kohta on näiteks ema täitnud püsielaniku ankeedi, õde kinnitab, et isik on välismaale lahkunud ja vend arvab, et isik elab Eestis ajutise elanikuna. Kõigi niisuguste olukordade puhuks tuli välja töötada vettpidavad otsustuseeskirjad, mis on kooskõlas rahvusvaheliste reeglite ja loenduste hea tavaga ja mis võimaldavad otsustada, kas isik on Eesti püsielanik või mitte. Asja teeb keerukamaks ka see, et rahvaloenduse puhul mõistetakse püsielanikku pisut teisiti kui tavaelus ja uuringutes. Näiteks on paljudel tudengivanematel raske omaks võtta, et rahvaloendusel on nende ülalpeetav laps omaette leibkond püsielukohaga ülikoolilinnas ja vanematekodu on tema jaoks vaid teine elukoht. Teine harjumatuna tunduv asi on rahvusvaheline kokkulepe selle kohta, et püsielaniku staatus tekib üldjuhul alles aasta jooksul: eelmisel kuul kodust lahkunu pole veel välismaalane.

Kõige olulisem ja kõige vanem tava, mida loendusandmete töötlemisel tuleb arvestada, on aga loendusandmete usaldamine. Loendusandmeid üldjuhul ei võrrelda teiste allikate (nt registrite) andmetega, et neid parandada ja täpsustada. Teisi andmeallikaid kasutatakse loendusandmete

täiendamiseks üksnes juhul, kui loendusandmetes on mõni näitaja puudu või selgelt vastuoluline (nn jäme viga).

Suurim on küsimustest arusaamise ja õige vastuse andmise probleem kindlasti iseloendamisel, mida Eestis põhilise meetodina (e-loendusel) kasutati esimest korda. Valikuliselt on seda ilmselt kasutatud ka varasematel Eesti Vabariigis toimunud loendustel (1922 ja 1934), kui loendajad võisid tuua peresse ankeedid, selgitada nende täitmist ja teisel külastusel ankeedid kokku korjata. Kahjuks puuduvad tollase isetäitmise mahu ja tulemuslikkuse kohta kokkuvõtted ja analüüsid.

Loendusele eelnenud teavitustöös selgitati põhjalikult, millal on inimene Eesti püsielanik ja millal välismaalane, vastav selgitus oli ka ankeetides, samuti teavitati põhjalikult loendajaid. Selletõttu ei kontrollita seda, kas isik, kes on enese alaliseks elukohaks märkinud Eesti (või kelle on Eesti elanikuks märkinud tema leibkonnaliikmed), tõepoolest Eestis elab. Kontrollimata jätmise põhjuseks on ka see, et loendamisele kuuluvad rahvusvaheliste reeglite kohaselt ka illegaalselt riigis viibivad isikud, keda ükski register ei kajasta. Küll aga tuleb analüüsida olukordi, kus loendusel laekunud info on vastuoluline.

Kui duplikaate on täitnud eri isikud, siis loetakse esmaseks (originaaliks) see ankeet, mille on täitnud isik ise või alla 15-aastase lapse vanem. Kui aga originaalis osa andmeid puudub, siis täiendatakse seda duplikaatide põhjal, kontrollides lisatavate andmete kooskõla originaalis olevate andmetega. Pärast esmaste ankeetide väljavalimist ja ülearuste duplikaatide kõrvaldamist on võimalik öelda loendatud isikute ja loenduse käigus selgunud riigist lahkunute arv. Selle töö tulemusena ongi juba selgunud esialgne loendatute arv. Nii loendatud isikute kui ka riigist lahkunute puhul tuleb aga arvesse võtta, et tegemist on loendusel laekunud infoga, mis paratamatult ei ole täiesti kõikne ja täpne.

Loendusandmete üle- ja alakaetus

Varasematel loendustel on loendatud isikute arvu käsitletud rahvaarvuna. Tegelikult ei ole see päris õige. Alati on jäänud osa inimesi mingil põhjusel loendamata, selle tagajärjel on loendustulemus tegelikust rahvaarvust väiksem. Niisugust olukorda nimetatakse alakaetuseks ja selle määraks on loendamata jäänud inimeste arvu suhe rahvaarvusse (väljendatud tavaliselt protsentides). Loendamisel saab tekkida teinegi viga – loendatakse liiga palju inimesi. Siis on tegemist ülekaetusega. Ülekaetuse määraks on liigselt loendatud inimeste arvu suhe rahvaarvusse. Ülekaetuse põhjuseks võib olla näiteks ajutiste elanike loendamine püsielanikena, kuid kõige sagedamini on põhjuseks isikute korduv loendamine. Ajalooliselt on rahvas olnud suhteliselt paikne, mistõttu ala- ning ülekaetus on olnud võrdlemisi väikesed ning enam-vähem tasakaalus. Loendusandmetest täpsemat tegeliku rahvaarvu hinnangut polnud kuskilt võtta, küll aga oli võimalik võrrelda rahvaloenduse andmeid varasemal rahvaloendusel saadud rahvaarvuga, mida oli täiendatud jooksva rahvastikustatistikaga. Seetõttu oligi rahvaloenduse tulemuse kasutamine rahvaarvuna õigustatud ja seda tehakse väga paljudes maades praegugi. Ka Eesti Statistikaameti avaldatava ametliku ja samuti rännet arvestava rahvaarvu aluseks on viimase rahvaloenduse andmed, mida aasta-aastalt muudetakse vastavalt sündide ja surmade arvule. Rändega rahvaarvu puhul arvestatakse ka sisse- ja väljarännet.

Tänapäeval on arenenud maades rahvaloenduste korraldus muutunud ja kvaliteedinõuded rangemaks läinud. Ülekaetust õnnestub isikukoode arvestades peaaegu täielikult vältida, alakaetus on aga kogu maailmas väga tõsine probleem. Inimesed on liikuvad, neil on mitu elu- ja töökohta ning loendajatel on raske neid kätte saada. Lisaks sellele varjab mingi hulk inimesi kiivalt oma isikuandmeid, lugedes nende avaldamist privaatsuse riiveks. On ka mitmesuguseid muid põhjuseid, miks osa inimesi jääb loendamata. Siit tulenebki tõdemus, et loendatud isikute arv ei ole tänapäeval täpselt sama, mis tegelik rahvaarv ja peamiseks loenduse probleemiks on alakaetus.

2000. aasta rahvaloenduse järel tehti järelloendus, et hinnata loenduse kvaliteeti, peamiselt kaetust. Ebu Tamme arvutuste kohaselt oli loenduse alakaetus 1,2%, kuid meedias väljendati ka arvamust, et alakaetus oli suurem. Eesti rahvaarvu loenduse alakaetuse arvel siiski ei parandatud, sest selleks puudus metoodika ja vajalik lisainfo vea jaotuse kohta soo-vanuse ja geograafilises lõikes.

2011. aasta rahva ja eluruumide loenduse alakaetus

Ka nüüdse loenduse järel on Eestis tulnud teateid loendamata jäänud inimestest. Esitame siin kaks tüüpilist meediaski kajastust leidnud näidet.

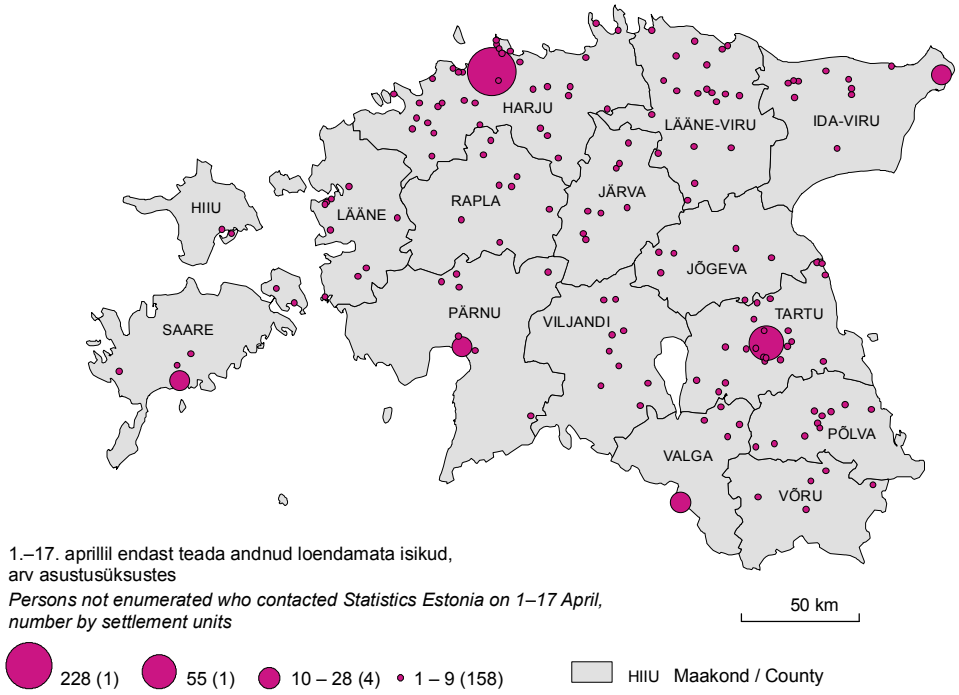
„Rahvaloendaja ei soovinud mind loendada, ütles, et minu korteris on juba kõik loendatud. Minu korterisse on sisse kirjutatud minu tütar, sest ma soovin, et tema päriks minu korteri. Tegelikult elab ta küll mujal, aga internetis loendades näitas, et elab siin.“ Tegemist on tüüpolukorraga, mida ka loenduse korraldajad ette nägid, kuid mille vastu on peaaegu võimatu rohtu leida. Kui inimesed loendasid end internetis valel aadressil (kõige sagedamini sellel aadressil, kuhu nad on ametlikult registreeritud) ja märkisid, et selles eluruumis on kõik isikud ja kõik leibkonnad loendatud, siis loendajad seda eluruumi vastavalt varasemale kokkuleppele ja elanikkonna teavitusele ei külastanud. Selles eluruumis tegelikult elavad inimesed, kes polnud end internetis loendanud, võisidki jääda loendamata. Toodud näite puhul on segaduses süüdi avalduse teinud mehe tütar, kes esiteks märkis enese eluruumi valesti, eirates kohustust märkida tegelik eluruum, ja kes teiseks, täites ankeeti isa eluruumi kohta, ei märkinud, et selles korteris elab ka tema isa.

Teine näide puudutab Tartu Supilinna. Eesti Raadio reporter, kes kontrollis rahvaloenduse usaldusväärsust, küsitles poe juures valikuliselt meesterahvaid ja avastas üsna mitu sellist, keda nende kinnitusel ei olnud loendatud. Näiteks üks küsitletud meesterahvas kõneles reporterile: „Mina ei ole loendajat näinud. Ta on küll minu maja juures käinud, aga ma elan teisel korrusel ja mul on uksed lukus. Postkasti oli ka sedeli jätnud, aga ma leidsin selle alles hiljem.“

Kummagi näite puhul pole alust süüdistada loenduse korraldust ega ka loendajate tegevust või tegevusetust, pigem on tegemist loendatavate ettearvamatu käitumisega.

Paratamatult on Eestis arvestatav hulk selliseidki inimesi, kes postkasti ei vaata, ust hoiavad lukus, telefoni ei kasuta ja kes selle tõttu on loendamata jäänud, kuigi nad on Eesti riigi püsielanikud. Nende „püüdmiseks“ pakkus loendusmeeskond veel mitut võimalust: loendaja sai telefoni teel koju kutsuda ning veel pärast loenduse lõppu sai endast märku anda loenduse kodulehel. Nii saadi teada veel mitmest tuhandest seni loendamata jäänud isikust (kaart 1).

Kaart 1. Loendamata jäänud isikud, kes endast pärast küsitluse lõppu teada andsid
Map 1. Persons not enumerated who informed about themselves after the Census



Kaardilt 1 on näha, et esialgu loendamata jäänud, kuid endast teada andnud isikute paigutus Eestis vastab ligikaudu rahvastikutihedusele: rohkem jäi inimesi loendamata suuremates linnades ja vähem hõredalt asustatud piirkondades. See näitab, et tegemist oli juhusliku protsessiga. Ometi ei jõudnud teave kõigi loendamata isikute kohta rahvaloendajateni: välja jäid niihästi teadlikult eemale hoidujad kui ka lihtsalt passiivsed või ebalojaalsed isikud. Võib oletada, et loendusandmed on vähemalt 1% ulatuses alakaetud, st, et nii palju inimesi on kõigist pingutustest hoolimata loendamata jäänud. Rahvusvahelises võrdluses pole see tulemus iseenesest halb, kuid kindlasti ei rahulda see teadmine kõiki neid inimesi, kes soovivad teada saada täpset Eesti rahvaarvu mitte ainult numbrina, vaid soo, vanuse ja rahvuse järgi jaotatuna koos täpse geograafilise paiknemisega.

Eestist lahkunud isikute loenduspõhised andmed

Täiendavat teavet tegeliku rahvaarvu kohta annavad loenduse käigus kogutud andmed Eestist lahkunud inimeste kohta. Kahjuks ei ole see teave täielik, sest kui mõni leibkond on täies koosseisus lahkunud, siis ei tarvitse ükski inimene nende kohta ka lahkunu ankeeti täita. Seega võib arvata, et loenduse käigus kogunenud lahkunute andmestik on veelgi enam alakaetud võrreldes loendatud Eesti elanike andmestikuga. Samas vajab see andmestik täpsustamist, sest ühelt poolt on lahkunute hulka märgitud inimesi, kes rahvusvaheliste reeglite kohaselt olid loendushetkel Eesti püsielanikud, sest lahkusid alles mõni kuu enne loendusmomenti või isegi pärast seda. Teiselt poolt on lahkunute andmestikus väga palju duplikaate, sest isiku võis lahkunuks märkida koguni mitu lähisugulast (vanemad, õed-vennad, abikaasa, lapsed).

Rahvastikuregister ja Eesti tegelik elanike arv

Järgmine tööetapp, mis REL-i meeskonnal ees seisab, ongi tegeliku Eesti elanike arvu hindamine. Selleks rakendatakse Eestis olemasolevaid registreid, kusjuures andmete sidumiseks kasutatakse krüptitud koode, mis isikute tuvastamist ei võimalda.

Kuigi Eesti registrid on valdavalt loodud ja käivitatud viimase kümmekonna aasta jooksul, mistõttu need on võrdlemisi noored ja nende kvaliteet pole kindlasti veel nii hea, et 2011. aasta loendust olnuks võimalik teha täies ulatuses registrite põhjal, sisaldavad need siiski üsna palju teavet, mida saab kasutada Eesti rahvastiku arvukuse hindamiseks.

Rahvastikuregister. Üks kõige vanemaid ja täpsemaid registreid Eestis on rahvastikuregister, mis põhimõtteliselt peaks kajastama kõiki Eesti elanikke koos nende põhiandmetega, mille hulgas on kõige olulisemad isikukood ja elukoht, sh elukoha riik. Eesti isikukoodiga inimesi, kes elavad siin ajutiselt või on siit lahkunud ja kelle elukohariik erineb Eestist, on rahvastikuregistris üle 83 000. Kahjuks on Eestis elavate isikute täpsed aadressiandmed rahvastikuregistri kõige nõrgem koht: elukoha registreerimine ei olnud Eestis rea aastate jooksul kohustuslik ja on praegugi kohustuslik üksnes formaalselt (seda üldiselt ei kontrollita). Tegeliku elukoha varjamiseks ja elukoha ekslikuks registreerimiseks on rida ajendeid: omavalitsuste soodustused, koolide jmt eelistused. Seetõttu ei vasta rahvastikuregistri elukoht umbkaudu viiendikul juhtudest inimese tegelikule elukohale. Kahjuks on nende juhtude hulgas ka sellised, kus inimene on Eestist lahkunud, kuid jäänud registris Eesti elanikuks. See põhjustab rahvastikuregistri ülekaetust, sest oletatavasti on sisserännanud registrisse kantud korrektsemalt (seda kinnitab ka rahvusvaheline praktika). See pole etteheide registrile, vaid pigem seadusandlusele. Siin avaldub Eesti elanike võrdlemisi nõrk registrikultuur ja lojaalsus, mis jääb oluliselt maha meile eeskujuks olevatest Põhjamaadest.

Võrreldes (krüptitud koodide alusel) rahvastikuregistri andmetel loendusmomendil Eestis elanud isikute loetelu loendatud isikute loeteluga, on võimalik leida loendamata jäänud isikud. Varasemates uuringutes toetust leidnud hüpotees rahvastikuregistri ülekaetusest sai kinnitust ka äsjasel loendusel: rahvastikuregistris Eesti elanikena märgitud inimeste arv (loendusmomendi seisuga 1 365 552 inimest) on suurem esialgsest püsielanikena loendatud isikute arvust 1 294 236. Rahvastikuregistrisse kuuluvate, ent loendamata isikute puhul on kaks võimalust: nad

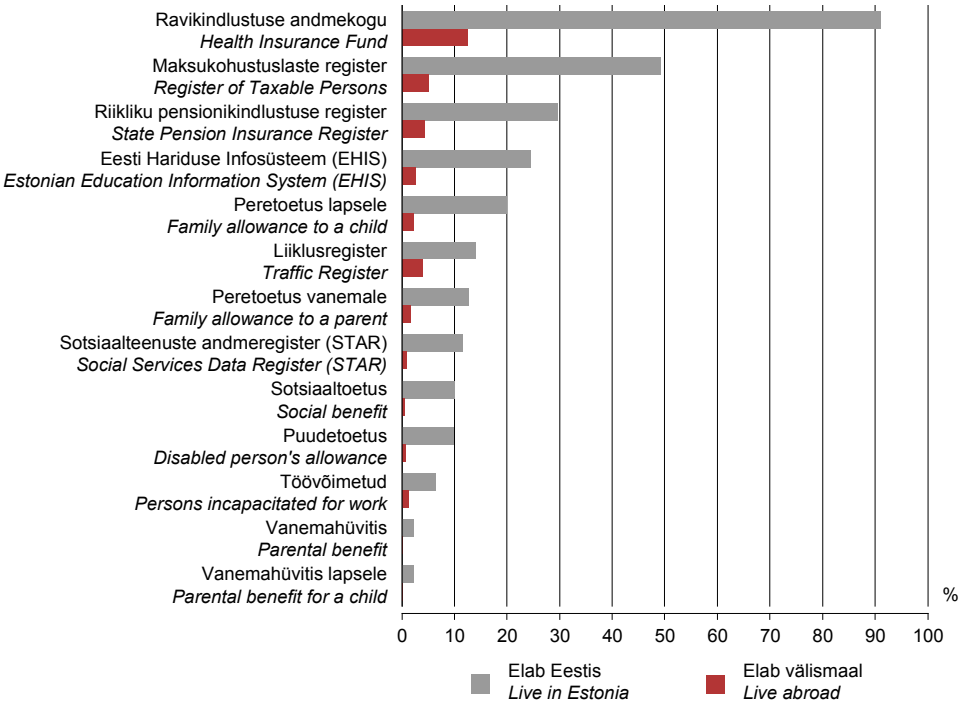
kas elavad Eestis ja jäid loendamata või nad on Eestist lahkunud, ilma et nad oleksid oma lahkumise registreerinud. Iga sellesse loetelusse kuuluva isiku kohta tuleb langetada otsus, kumba kategooriasse ta kuulub: kas ta on Eesti püsielanik või on Eestist lahkunud. Nimelt selleks tulebki kasutada teisi registreid.

Teised Eesti registrid ja nende kasutamine tegeliku elanike arvu määramiseks

Nende isikute kohta, keda loendusel ei loendatud ei püsielanikuks ega ka lahkunuks, tuleb leida täiendava info allikad. Selleks saab kasutada paljusid teisi Eestis toimivaid registreid (vt joonis 1). Nende seas on olulisemad **ravikindlustuse andmekogu**, mis sisaldab üle 90% Eesti inimestest, **maksukohustuslaste register**, mis katab olulise osa tööealisest elanikkonnast, **Eesti Hariduse Infosüsteem (EHIS),** mis sisaldab peaaegu kõikselt õppijate andmeid ja on seega väga hea allikas koolieaste isikute tuvastamiseks. Üsna palju täiendavat infot on võimalik saada veel **riiklikust pensionikindlustuse registrist** (sisaldab nii pensionide, peretoetuste kui ka hüvitiste saajaid), **sotsiaalteenuste ja -toetuste andmeregistrist** (**STAR**), samuti **liiklusregistrist**, mis sisaldab niihästi sõiduki kasutajate kui ka omanike andmeid. Kuna vanemahüvitise ja peretoetusega on seotud nii vanem kui ka laps(ed), siis on neid mõlemaid registrisse kantuna arvestatud.

Joonis 1. Eestis aktiivselt tegutsenud inimeste osatähtsus Eesti registrites rahvastikuregistri andmetel alaliselt Eestis ja välismaal elavate inimeste hulgas, 2011

Figure 1. Share of persons having actively operated in Estonia in Estonian registers among persons residing permanently in Estonia and abroad according to the data of the Population Register, 2011

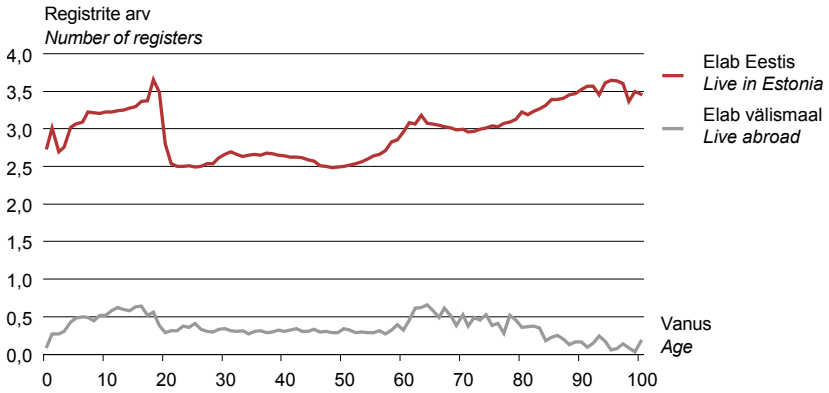


Vaadates kõigis nimetatud registrites 2011. aastal aktiivselt toiminguid teinud isikuid, selgub, et keskmiselt on iga Eesti elanik aasta jooksul tuvastatud (peale rahvastikuregistri) veel kolmes registris (vt joonis 2). Kuid on ka isikuid, kes on küll rahvastikuregistri andmetel märgitud Eesti

elanikuks, kuid kelle kohta aasta jooksul üheski muus registris ühtegi märget pole tehtud. Loogiline oleks arvata, et niisugune inimene on Eestist lahkunud. Niisugust otsust ei saa siiski langetada ülepeakaela, eelnevalt tuleb analüüsida kõiki registreid ja elanikkonna kaetust nende registritega. Alljärgnevas esitamegi vastava analüüsi skeemi.

Joonis 2. Eri vanuses isikute keskmine esindatus registrites rahvastikuregistri andmetel alaliselt Eestis ja välismaal elavate inimeste hulgas, 2011

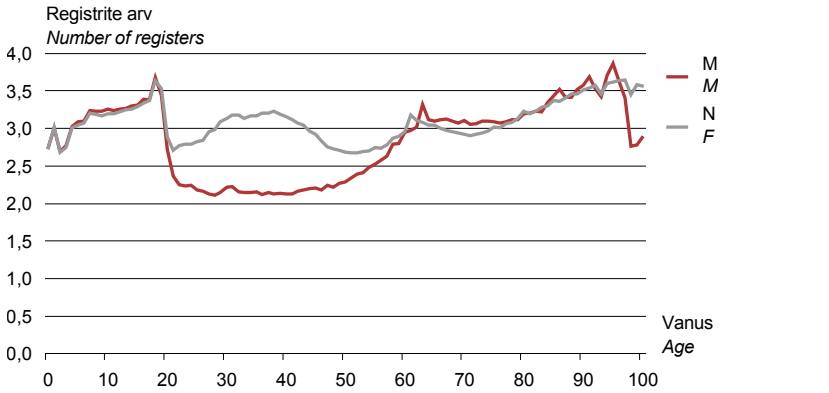
Figure 2. Average number of persons of different age in the registers among persons residing permanently in Estonia and abroad according to the data of the Population Register, 2011



Inimesed, kes on peale rahvastikuregistri veel mitmes muus registris aktiivsete tegutsejatena kirjas, elavad suure tõenäosusega Eestis ja kui nad ei ole loendatud, võib selle põhjuseks olla see, et nende aadressil loendati kedagi teist (kui viiendik kõigist inimestest, kelle tegelik ja registreeritud elukoht erinevad, märkis e-loendusel oma elukohaks registreeritud eluaseme, võis see põhjustada ligi 1% inimeste loendamata jäämist). Loendamata jäämise põhjuseks võis olla ka ajutine välismaal viibimine või lihtsalt kõrvalehoidumine. Otsust, et inimene elab Eestis, aga jäi loendamata, ei saa üldjuhul langetada ainult ühe registri põhjal, sest küllap on ka selliseid inimesi, kes kasutavad mõningaid Eesti sotsiaalteenuseid, olles ise välismaale elama asunud. Jooniselt 2 selgub, et registritega kaetus sõltub üsna oluliselt isiku vanusest ja kõige vähem on registritega kaetud tööealised 20–60-aastased isikud, eriti mehed (vt ka joonis 3).

Joonis 3. Rahvastikuregistri andmetel Eestis elavate meeste ja naiste keskmine esindatus teistes registrites, 2011

Figure 3. Average number of men and women residing in Estonia in other registers according to the data of the Population Register, 2011



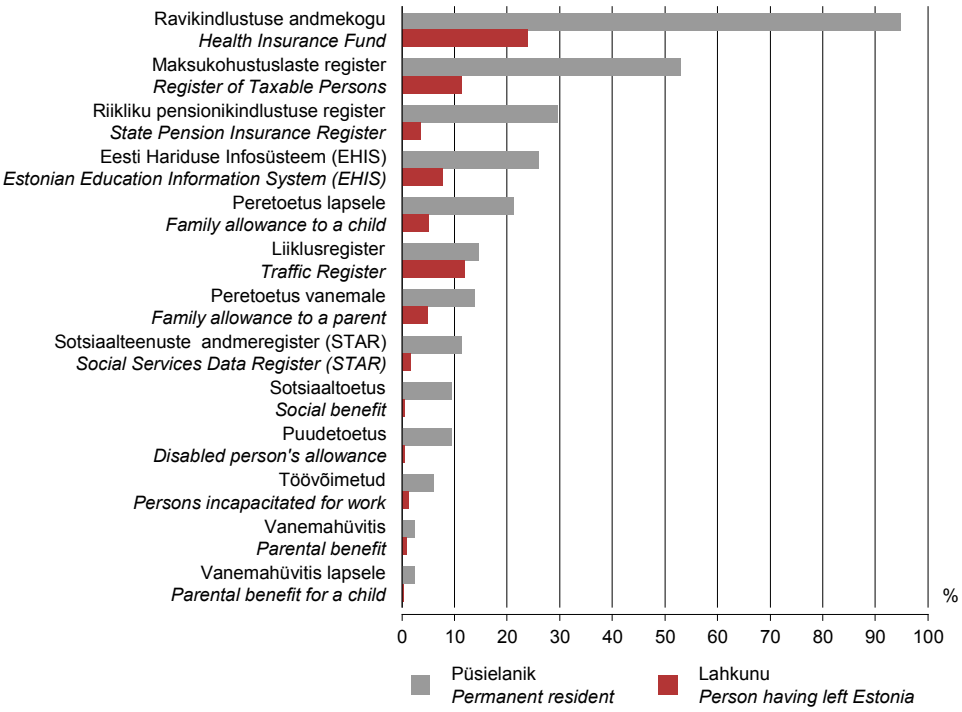
Loendusandmete võrdlus registriandmetega

Märtsi alguses, kui loendusinfot oli kogunenud juba rohkem kui miljoni inimese kohta, kelle hulgas oli niihästi püsielanikke kui ka lahkunuid, võrreldi kahe loendatud isikute kategooria – püsielanike ja lahkunute – aktiivsust 2011. aastal registrite andmetel (vt joonis 4).

Joonistel 1 ja 4 esitatud andmeid kõrvutades selgub, et loendatud püsielanike hulgas on rahvastikuregistris olevate Eesti elanikega võrreldes suhteliselt rohkem neid, kes on aktiivsed niihästi maksukohustuslaste, EHIS-e, haigekassa, liikluskindlustuse kui ka vanemahüvitise ja peretoetuste mõttes. Kui aga võrrelda rahvastikuregistri andmetel välismaal elavate inimeste ning REL-i andmete põhjal Eestist lahkunute aktiivsust registrite alusel, siis oli viimane enamiku registrite puhul märgatavalt kõrgem.

Joonis 4. Eestis aktiivselt tegutsenud inimeste osatähtsus Eesti registrites REL2011 käigus kuni 10. märtsini loendatud püsielanike ja lahkunute hulgas, 2011

Figure 4. Share of persons having actively operated in Estonia in Estonian registers among persons residing permanently in Estonia and having left Estonia who have been enumerated until 10 March during PHC2011, 2011



Selle andmestiku põhjal töötatigi välja esmane eeskiri rahvastikuregistrisse kuuluvate, kuid loendamata jäänud isikute jaotamiseks kahte kategooriasse: Eesti püsielanikud ja mitteametlikult lahkunud. Ülesanne osutus keeruliseks, sest registrites sisalduv info on kohati vastuoluline, mida osaliselt põhjustab inimeste teadlik käitumine. Siiski selgus rida usaldusväärseid registritesse kuulumise kombinatsioone, mis väga suure tõenäosusega (väga väikese eksimisriskiga) lubavad otsustada selle üle, et inimene tõepoolest Eestis elab (tabel 1).

Tabel 1. Registrite kombinatsioonid ja vastav riskisuhe isiku Eestis elamise üle otsustamiseks

Table 1. Combinations of registers and the corresponding risk ratio for determining the person's residence status in Estonia

Ravi-kindlustuse andmekogu *Health Insurance Database*	EHIS[a] *EHIS[a]*	Maksu-kohustuslaste Register *Register of Taxable Persons*	Riiklik pensionikindlustuse register *State Pension Insurance Register*	STAR[b] *STAR[b]*	Liiklusregister *Traffic Register*	Riskisuhe *Risk ratio*
	+		+	+		19,65
			+	+		12,14
	+		+			9,39
+		+	+	+	+	7,40
+		+	+	+		6,82
+			+	+		6,13
		+	+	+		5,92
+		+	+			4,86
+		+	+		+	4,52
+		+		+	+	3,90

[a] EHIS – Eesti hariduse Infosüsteem
[b] STAR – sotsiaalteenuste andmeregister

[a] EHIS – Estonian Education Information System
[b] STAR – Social Services Data Register

Seda, et inimene ei ela Eestis, on märksa raskem veenvalt tõestada. Püsielanikena loendatud Eesti elanike seas on üle 2% isikuid, keda ei näita ükski register, neile lisandub ligi 4% inimesi, kes kajastuvad vaid ühes registris. Pole võimalik lähtuda eeskirjast, mis kõik niisugused inimesed välismaalasteks liigitaks. Selgub, et ka rahvastikuregistris olevate Eesti elanike seas on üsna palju inimesi, keda ükski teine register ei kajasta või kes kajastuvad ühes registris, kusjuures suur osa neist on püsielanikena loendatud ja pole vähimatki alust kahtlustada neid selles, et nad Eestis ei ela. Seega tuleb leida veel täiendavaid infoallikaid selleks, et täiustada loendamata jäänud Eesti elanike arvukuse hinnangut. See töö jätkub.

Eesti rahvaarvu ja rändesaldo hindamine

Eesti rahvaarvu määrataksegi, lisades loendatud püsielanike arvule nende isikute arvu, kes jäid loendamata, kuid arvestades muud infot (toimivad registrid) elavad väga suure tõenäosusega püsivalt Eestis. Niiviisi arvutatud rahvaarvu põhjal on võimalik hinnata ka REL2011 kaetust.

Suurt huvi pakub ka loendustevahelisel perioodil Eestist lahkunud inimeste arv. Registreeritud rände saldo selgub ametliku rännet mittearvestava ja rännet arvestava rahvaarvu vahest, mis on 21 653 inimest. Mitteametliku (registreerimata) rände hinnang koosneb kahest komponendist. Ühe neist saame, kasutades rahvaloendusel saadud infot lahkunute kohta, võttes nende seast arvesse ainult need, kes rahvastikuregistri andmetel praegu Eestis elavad. Teise osa mitteametlikust rändest moodustavad rahvastikuregistri andmetel Eestis elavad loendamata isikud, kes registrite analüüsi alusel liigitati lahkunute hulka. Võrreldes rahvastikuregistri andmeid loendatute arvuga, saame hinnata ka mitteametlikku sisserännet, seega tekib lisaks registreeritud rändesaldo arvule ka registreerimata rände saldo hinnang.

Paratamatult jääb alles teatav eksimise võimalus. Võib-olla jääb kõigest hoolimata Eesti püsielanike loetelust välja näiteks meesterahvas Supilinnast, kes elatub pudelite korjamisest ja sotsiaalabi ei vaja, pensioniikka pole veel jõudnud ning pole olnud juhust ka viimase aasta jooksul arsti juurde sattuda, mistõttu ei ole tema tegevusest viimase aasta jooksul üheski registris ühtegi jälge. Loendusmeeskond püüab teha parima, et saada kogu olemasolevat infot kasutades võimalikult täpne elanike arvu hinnang. Statistikute jaoks ei ole parim mitte „võimalikult suur", vaid „võimalikult täpne".