

Statistikanõukogu, 26. november 2019

Eesti 2021 rahvaloendus: loendusmeetodi valikust

Allan Puur

Tallinna Ülikool, ühiskonnateaduste instituut



TALLINNA ÜLIKOOL
Eesti Demograafia Instituut

Loendusmeetodi valik: lähtekohad

- ▶ 2021. aasta loendusel on valida **kahe variandi** vahel:
 - **täielikult (100%)** registripõhine loendus
 - **osaliselt (80+%)** registripõhine e. kombineeritud loendus
- ▶ Tehtud **eeltöö** on **põhjalik** ja **tulemuslik**
 - 2011 vaid **üks** (osa)tunnus, mis võeti registritest
 - 2021 **enamuse** loendustunnustest saab registritest
- ▶ **Kuid on kaks probleemi**, mida valikul arvestada
 1. **Elukohaandmete ebatäpsus**
(20-25% isikute registrijärgne ja tegelik elukoht lahknevad, mis moonutab kuni 50% pere- ja leibkondade andmeid)
 2. **Mitme olulise** loendustunnuse **puudumine** registrites
(sh eesti keele oskus, võõrkeelte oskus, tegevuspiirangud)



Kas (1.) probleem eksisteerib ka teistes registripõhist loendust kasutavates maades?

Riik	Tegeliku ja registripõhise elukoha lahknevusmäär (% rahvastikust)
Soome	1,9%
Holland	2,5%
Norra	5%
Rootsi	Sama suurusjärg kui teised Põhjamaad
Eesti	20–25%

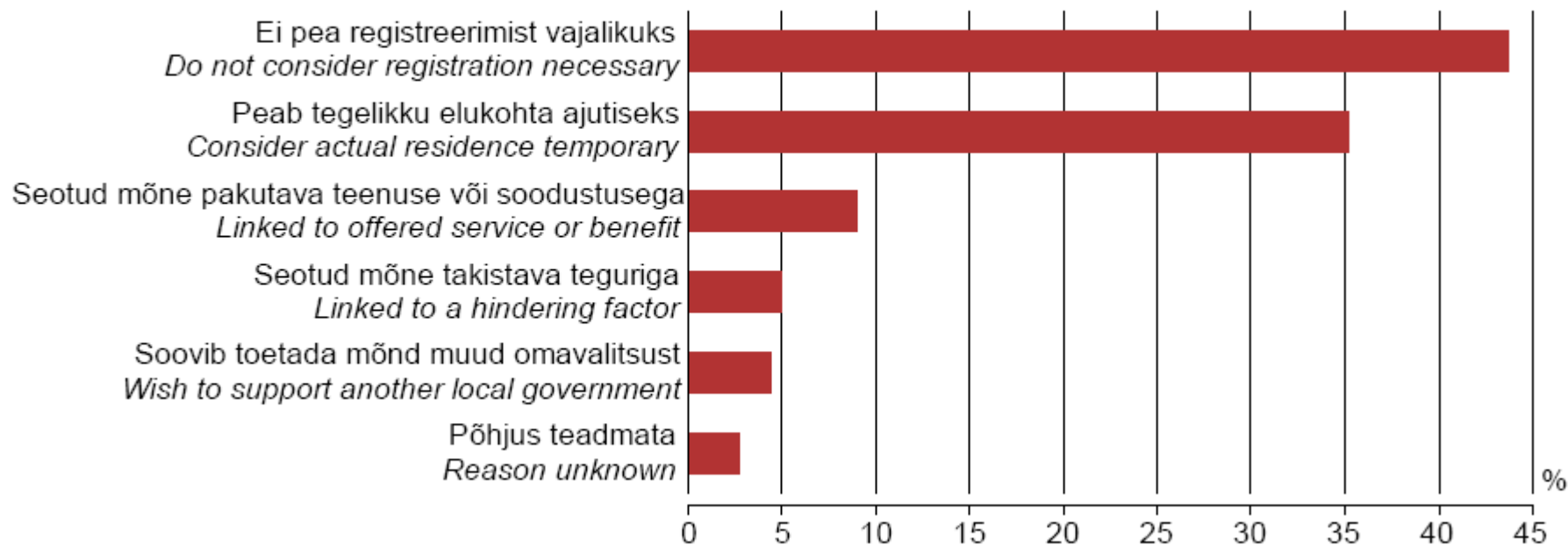
Allikad: Soome statistikaamet (jari.nieminen@stat.ee), Hollandi statistikaamet (cn.harmsen@cbs.nl)
Norra statistikaamet (kva@ssb.no), Rootsi statistikaamet (kristian.soderholm@scb.se),
Eesti staistikaamet (<https://blog.stat.ee/2019/11/08/regrel-kas-lugeda-rahvast-savitahvlite-voi-taimejuurte-abil/>)



Mis on elukohainfo suure lahknevuse põhjus?

Joonis 1. Tegelik ja registreeritud elukohta erinevuse peamine põhjus, 2015

Figure 1. Main reason for difference in actual and registered place of residence, 2015



Allikas: Helerin Äär . Registripõhise ja tegeliku elukoha kattuvus. *Eesti Statistika Kvartalikirj* 1/2017, lk . 73–79.

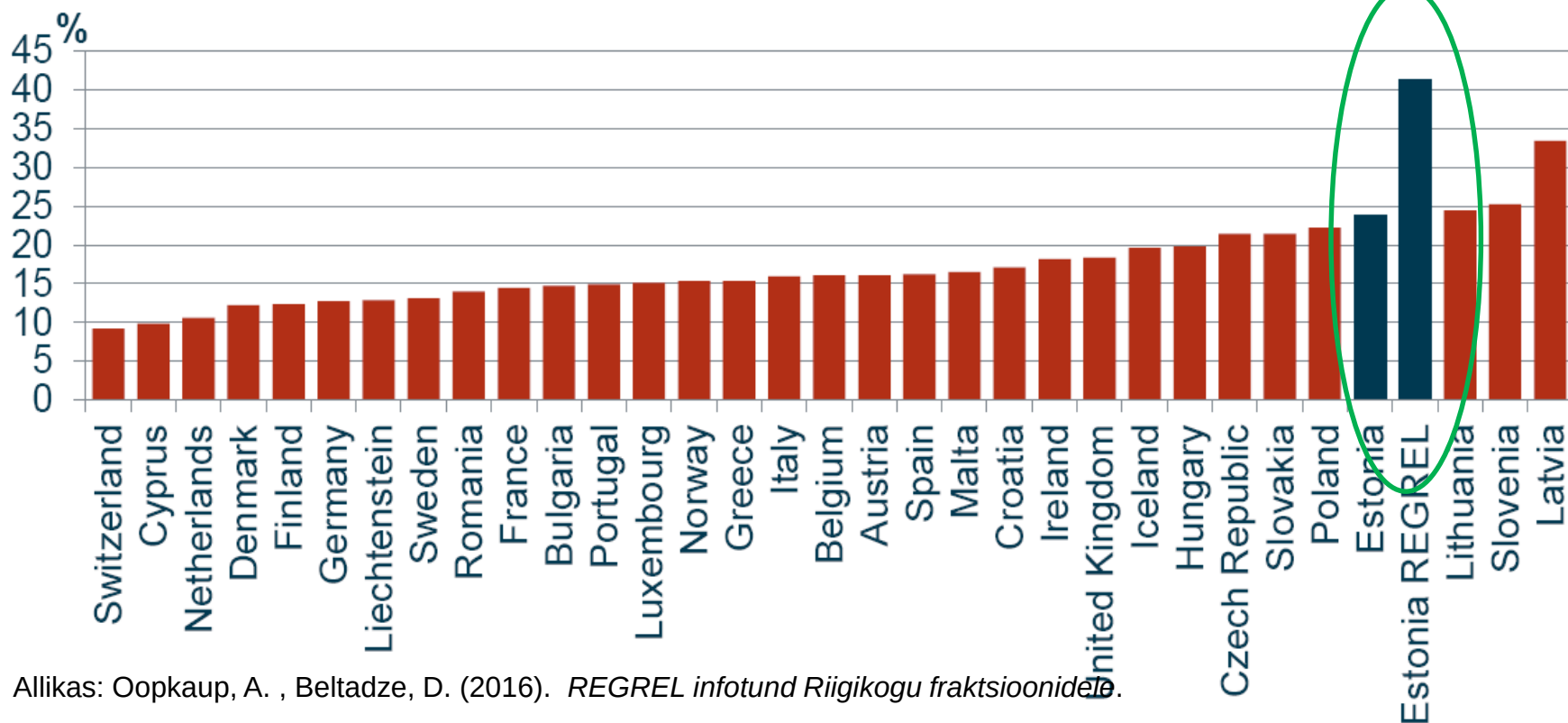


Mis on elukohainfo suure lahknevuse tagajärg?

[2016: 1. prooviloendus]



Üksikvanema pere osatähtsus



Allikas: Oopkaup, A. , Beltadze, D. (2016). *REGREL infotund Riigikogu fraktsioonidele*.



TALLINNA ÜLIKOOL
Eesti Demograafia Instituut

Väljapakutud lahendus

- ▶ Kõik olid nõus, et rahvastikuregistrist **ei ole võimalik saada** piisavalt **usaldusväärset teavet** elukoha, leibkonna ja perekonna kohta
- ▶ **Pakutud lahendus: mudelarvutus** (indeksimeetod)
 - 1/ Nähtuste kohta leiduvad registrites **kaudsed „märgid“** (nt kooselupartneriga ühised lapsed, kinnisvara, auto, kasutus, ühine tuludeklaratsioon jms)
 - 2/ „**Märgid**“ pole nähtustega üksüheses, vaid **tõenäosuslikes seostes** (need tuvastatakse matemaatiliste mudelitega)
 - 3/ „Märkide“ alusel ja mudeli abil **arvutatakse** isiku **tõenäoline seisund või suhted** (nt tõenäoline kooselupartner)
- ▶ **3 indeksit:** residentsus-, partnerlus- ja paiksusindeks



Lahenduse kriitika

- ▶ **Kontseptuaalne kriitika:** mudel on alati **lihtsustus (prognoos)**, loendus on aga **statistiline vaatlus**, mille roll on tagada võimalikult detailitäpne tegelikkuse „jäädvustus“.

„Arusaadavalt on andmestik üldiselt seda parem, mida rohkem tunnuseid on mõõdetud vahetult ja mida vähem imputeeritud“ – Ene-Margit Tiit (2011. aasta rahva- ja eluruumide loendus. Metoodika“, lk. 8.

- ▶ **Empiiriline kriitika:** lihtsustuse tõttu mudelarvutus:
 - tekitab andmetes **süsteemaatilisi moonutusi**
 - loendusandmete **aegread** võivad **katkeda**
- ▶ **Taustaks:** Registripõhise statistika **eeskujumaad mudelarvutust** SA poolt pakutud viisil/ulatuses **ei kasuta:**

„Finland (as well as other Nordic countries) has not used mathematical (probabilistic) models to obtain the values of census variables for individuals. And there isn't any change with this regard from the 2010 to 2020 census round“ – jari.nieminen@stat.fi, 22.11.2019



Lahenduse kriitika(2): mudeli (partnerlusindeks) abil arvutatud kooselupaaride valideerimine

Vanusrühm (naised)	Tõenäosus, et mudeliga leiti tegelik paar (ESU ja ESU võrdlusandmed)	Tõenäosus, et mudeliga leiti tegelik paar (LEKU võrdlusandmed)	
Alla 30- aastased	69%	63%	suur süste- maatiline moonutus
30–59-aastased	88%	87%	
60+ aastased	92%	91%	
Kokku	87%	84%	suur ebatäpsus

Mudel on arvesse võetud 16 elumärki: registreeritud elukoht, sõiduki ühiskasutus, ühine laps, abielu, lahutus, ühine kinnisvara, ühine eluasemelaen, digiretsept, isapuhkuse kasutamine, maksuvaba tulu ülekandmine, hooldusleht, vanemahüvitis, toimetulekutoetus, elatis jms



Mida ütleb selle probleemi kohta

2. prooviloenduse raport?

Vastus: mitte midagi uut.

Miks? SA ei pidanud vajalikuks seda põhjalikumalt analüüsida.
(SA hinnangul piisab, kui tunnuste üldjaotused on realistlikud, vastav kvaliteedikriteerium on nii seatud)

Prooviloenduse hindamise kriteeriumid (väljavõte):

Hinnangu andmisel arvestatakse järgmisi kvaliteedidimensioone ja -näitajaid:

Kvaliteedi-dimensioon	Kvaliteedinäitaja	Arvutuseeskiri	Hinnang
Seostatavus	Kokkulangevuse määr võrdlusallikaga	Kokkulangevuse määr leitakse isiku-uuringute baasil moodustatud võrdlusbaasi andmete põhjal. /.../ Koondandmete tase: absoluutkeskmise jaotuste erinevus protsendipunktides	5 – absoluutkeskmise erinevus <3,0; 3 – absoluutkeskmise erinevus 3,0– 5,0; 1 – absoluutkeskmise erinevus >5,0.



ÜRO lähenemine loenduse andmekvaliteedi analüüsile



United Nations Statistics Division

Overview of Census Evaluation Methods

United Nations Statistics Division

UN Regional Workshop on Evaluation and Analysis of Census Data by UNFPA and UNSD,
Johannesburg, South Africa, 18-22 September 2017



TALLINNA ÜLIKOOL
Eesti Demograafia Instituut



What are census errors?

☐ Coverage errors:

- Errors in the count of persons or housing units resulting from cases having been “missed” or counted erroneously

☐ Content errors:

- Errors in the recorded characteristics of persons or housing units resulting from the interview (enumerators/respondents), coding, editing, etc.

Arvesse lähevad kõik sisuved

UN Regional Workshop on Evaluation and Analysis of Census Data by UNFPA and UNSD,
Johannesburg, South Africa, 18-22 September 2017





Methods for the evaluation of census errors

Single Source of Data (rely only on the census being evaluated)

- ❑ Demographic analysis
 - *Empirical Consistency checks*
 - *Comparison with theoretical models*

Multiple Sources of Data

- ❑ Non-matching studies
 - *Demographic analysis using multiple census rounds*
 - *Comparison with administrative sources or existing surveys*
 - ❑ Matching studies ←
 - *Post Enumeration Surveys*
 - *Record checks*
- Välite andmeallikatega linkimine
ja isikupõhine kirjete võrdlus on
loenduse kvaliteedihindamise
standardmeetod

UN Regional Workshop on Evaluation and Analysis of Census Data by UNFPA and UNSD,
Johannesburg, South Africa, 18-22 September 2017



Pakutud lahendus 2. probleemile

- ▶ Registrites **puuduvate tunnuste asendamine valimuuringust** saadud tunnustega

- ▶ **Lahenduse kriitika:**

- **kontseptuaalne kriitika:** valimipõhine andmestik ei saa täisväärtuslikult asendada kõikset andmestikku

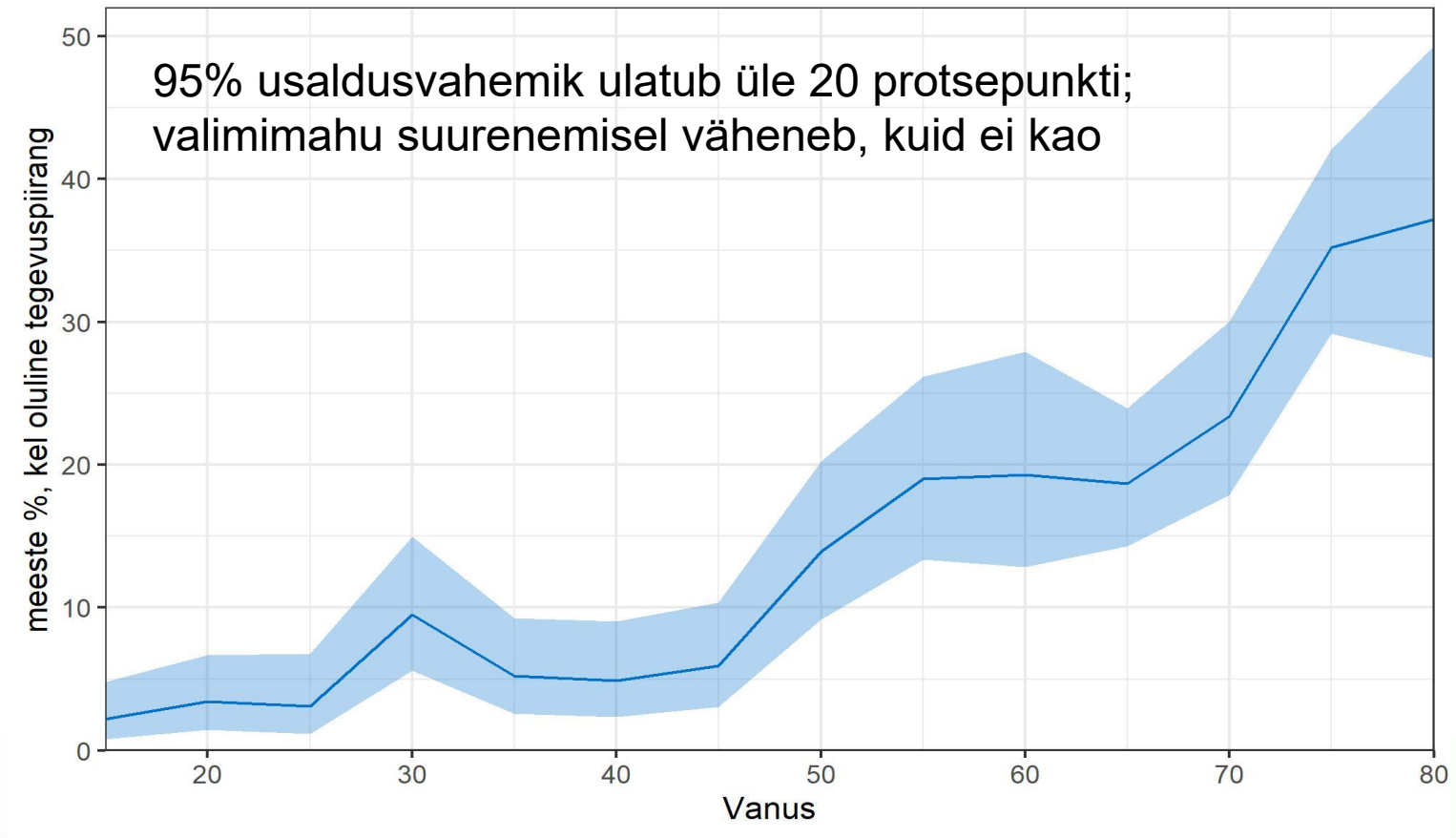
„Loenduse põhimõtted: rahva- ja eluruumide loendus peab koguma andmeid *kõigi* riigis elavate/paiknevate inimeste ja eluruumide kohta; loenduse käigus tuleb koguda andmeid *ka väikeste* geograafiliste piirkondade ja rahvastikurühmade kohta“, Ene-Margit Tiit, ibid. lk.4.

- valimuuring **ei taga piisava täpsusega** andmeid **väikerühmade ja piirkondade** kohta



Lahenduse kriitika (2): esindusviga

[Olulise tervisepiiranguga meeste osakaal,%]



Allikas: Eesti terviseuuring, autori arvutus



TALLINNA ÜLIKOOL
Eesti Demograafia Instituut

Kokkuvõtteks

- ▶ Soovitada Eesti 2021 rahvaloendus teostada **osaliselt registripõhisena**, kogudes **väikese arvu oluliste tunnuste** kohta (tegelik elukoht, leibkond, ...) teavet isikutelt
 - ▶ Tegemist on kompromissiga, mis aktsepteerib osalist kvaliteedi halvenemist ja võrraldamatust, kuid seab sellele siiski teatud piirid.
 - ▶ Isikutelt info kogumiseks kasutada **maksimaalselt e-loenduse võimalusi**, tuginedes 2011. aasta kogemusele.
- ▶ **Jätkata** jõuliselt administratiivregistrite **arendamist** ja **rakendamist** statistikas, sh tegelda tõsisemalt elukohainfo täpsuse parandamisega ja metoodikate täiustamisega
 - ▶ Lisaks SA töödele algas spetsiaalne RITA-rakendusuuuring Siseministeeriumi tellimusel (2019–2021).
- ▶ Täiendada **(proovi)loenduse hindamise metoodikat**
 - ▶ Loendus peab olema kvaliteetne ka rahvastikurühmade tasemel ega tohi sisaldada suuri süstemaatilisi moonutusi, oluline on analüüsida kvaliteeti ka isikupõhiste võrdluste kaudu (ÜRO käsitus)



Aitäh kuulamast!

nature > editorials > article

a nature research journal

MENU

nature

Subscribe



EDITORIAL • 16 OCTOBER 2019

Protect the census

The humble census risks becoming a casualty of the rush to embrace big data.
But it has the potential to save lives.



PDF version



Get the most important science stories of the day, free in your inbox.

Sign up for Nature Briefing



TALLINNA ÜLIKOOL
Eesti Demograafia Instituut

Mõned lisaslaidid

Vastuargumendid kombineeritud
loendusele ja nende kriitika



Levinud vastuargumendid kombineeritud loendusele (1)

► Küsitluspõhisel loendusel inimesed luiskavad

„Kuigi loogika võib öelda, et inimesed on rahvaloendajatega näost-näku suheldes palju ausamad kui ametlikke dokumente täites, siis tegelikult tekib nii hoopis rohkem infomüra“ – Diana Beltadze, usutluses Õhtulehele 13.09.2019

► Argumendi kriitika:

- **igas andmestikus leidub vigu**; küsimus on kui palju neid/kas nad on juhuslikud või süstemaatilised;
- **teaduskasutus ei kinnita**, et loenduste andmed oleksid tugevasti ja süstemaatiliselt moondunud;
- mudelarvutuse **tõe kriteeriumina** kasutatakse just inimestelt kogutud teavet.



Levinud vastuargumendid kombineeritud loendusele (2)

► Veebiloendusel antakse vähem täpset infot

SA korraldatud aruteludel on väidetud, et veebiloenduse tulemused on oluliselt kehvemad kui näost-näku küsitlusloendusel.

► Argumendi kriitika:

- „Enamiku võrreldavate küsimuste puhul oli **internetis** protsentuaalselt **vähem neid**, kes **vastamata jätsid** või **vastamistest keeldusid**“, Ene-Margit Tiit, ibid, lk 68.
- „Üllatav on see, et **aadressivigu** esines **küsitlusloenduse ankeetides rohkem** kui e-loenduse ankeetides“, Ene-Margit Tiit, ibid, lk 69).
- 2011. a. loenduse aadressandmeid **ei ole** tegelikult rahvastikuregistri aadressandmetega **terviklikult võrreldud**.



Levinud vastuargumendid kombineeritud loendusele (3)

► Nagunii tuleb üle minna täielikult registripõhisele loendusele

„Alates 2024. aastast peavad kõik Euroopa Liidu liikmesriigid esitama Brüsselile igal aastal loendusandmeid. Selle jaoks vajaminev info tuleb info tuleb kindlasti registritest“ — Õhtuleht 13.09.2019, usutlus Mart Mägi ja Diana Beltadzega

► Argumendi kriitika:

- Iga-aastaselt nõutav tunnuste hulk on **väiksem kui loendusel**;
- Juba praegu avaldab Eurostat riikidelt saadud iga-aastasi rahvastikukoostise andmeid **7 tunnuse lõikes**;
- **Registriandmete kasutamine** iga-aastaste seisude andmisel ongi **mõistlik**, kuid põhjalikuma info kogumine üks kord kümnendi vältel on teine ülesanne.



Levinud vastuargumendid kombineeritud loendusele (4)

- ▶ Kombineeritud loenduse **andmed jääksid vaid Statistikaametisse; otsused** tehakse nagunii vaid registriandmete põhjal
- ▶ **Argumendi kriitika:**
 - Statistikaamet **ei edasta ka täielikult registripõhiselt loodud andmeid** ametkondadele haldusotstarbel kasutamiseks; ka **indeksite abil loodud loendusandmed erinevad** rahvastikuregistri andmetest;
 - Statistikaameti poolt seatud tingimustel **saavad** ametkondade analüütikud **loenduse andmeid** (ka anonümiseeritud individuaalandmeid) **kasutada**;
 - Infoühiskonnas on **erinevate andmete olemasolu** ja kasutamine **pigem reegel kui erand** (nt töötus, tervisenäitajad jne); andmete mitmekesisus pole probleem.



Küsitavused prooviloenduse hindamiskriteeriumides

- ▶ Neljale hindamiskriteeriumile on antud võrdne kaal, mistõttu **olulisemat ei eristata vähem olulisest**
- ▶ Võrreldavust teiste allikatega hinnatakse vaid üldkeskmiste, tasemel mistõttu **süsteemaatilisi moonutusi rühmade** (nt vanusrühmad, sisserännanud) **tasemel eiratakse**
- ▶ Võrreldavust teiste allikatega hinnatakse vaid agregeeritud tasemel, mistõttu **üksikisiku tasandi vigade ulatust eiratakse**
- ▶ Koondhinnangud saadakse üldjuhul aritmeetilise keskmise põhimõttel, mistõttu osa tunnuste kasina kvaliteedi tasakaalustab teiste parem kvaliteet. **Nõrgima lüli põhimõtet eiratakse.**

