

Andmeteadus eluteaduste teenistuses

Eesti StatistikaSeltsi 29. konverents

23.-24.03.2018, Narva, TÜ Narva Kolledž

Valik ettekandeid

Konverentsi programm

23.märts 2018

13.00 Avasõnad, Kaja Sõstra, ESS president

13.10 Tervitus Narva linna poolt – Katri Raik, Narva linna volikogu esindaja

13.20-15.00 I sessioon: Tervis ja tervisekäitumine, juhatab Tõnu Kollo

1. Krista Fischer – Eluea ja suremuse geneetika ja epidemioloogia: kas pikk eluiga on saatus või valik?
2. Nele Taba – Mida ütleb Sinu kohta see, mida Sa sööd. TÜ EGV kohordi põhjal tuletatud toitumismustrid ning nende seosed tervisega
3. Kristi Läll – Levinud komplekshaiguste geneetiline risk – mis see on ja kuidas me saame seda kasutada?
4. Gudrun Veldre, Tanel Kaart – Müokardiinfarkti registrist
5. Mait Raag – Süstivate narkomaanide arv Eestis

15.00-15.30 kohvipaus

15.30-16.30 Tarmo Soomere (TA president) – Statistika ja dünaamika sümbioos mere- ja rannikuteaduses

16.30-17.50. II sessioon: Bioloogia ja metsandus, juhatab Kalev Pärna

1. Urmas Kõljalg – Elurikkusest
2. Ants Kaasik – Kodukootud fülogeneetiline meta-analüüsi mudel
3. Andres Kiviste ja Ando Lilleleht – Puistu struktuuri kirjeldamisest metsa proovitükkide mõõtmisandmeil
4. Mait Lang ja Tauri Arumäe – Aerolaserskanneerimise andmete ja puistu struktuuri seosed - kas andmeid on liiga palju või liiga vähe?

18.00 Õhtune programm (giidiga jalutuskäik Narvas, ühine õhtusöök)

24.märts 2018

8.30-9.00. ESS üldkoosolek

9.00-11.00 III sessioon: Vaade ajalukku läbi statistika, juhatab Tiit Paas

1. Mati Rahu – Eesti Vähiregister 50 (40min)
2. Lehti Saag – Eesti demograafilise ajaloo uurimine vana DNA abil
3. Peeter Hõrak – Menarche vanus ja elutempo evolutsioon
4. Ene-Margit Tiit – Sündimus ja selle mõõtmine
5. Mihkel Servinski – Suremus Eestis 100 aasta jooksul

11.00-11.20 kohvipaus

11.20-13.20 IV sessioon: Eluteaduste erinevad tahud läbi statistika, juhatab Imbi Traat

1. Shinya Sugita – Paleoecology (ingl. keeles)
2. Tatjana von Rosen – Lineaarsete mudelite rakendamisest eluteadustes
3. Märt Möls, Kristel Luik – Statistiku geeniandmete keskel
4. Meelis Kull – Tõenäosuste kalibreerimine klassifitseerimisel - paraja enesekindluse heaks
5. Liis Starkopf – Mediation analysis: vahend põhjuslike mehhanismide uurimiseks
6. Jaak Sõnajalg – Statistiku rollist ravimiuuringutes

13.20-14.00 lõuna

14.00-15.00 ümarlaud (moderaator SA peadirektor Mart Mägi) Kui kaugele suudab ja võib andmeteadus minna eluteaduse mudelites

Sisukord

Antud kogumikus publitseeritud ettekanded

Jaak Sõnajalg. Statistiku rollist ravimiuuringutes

Ene-Margit Tiit. Sünnimusest, sündimuse mõõtmisest ja sündimuse mudelistest

Gudrun Veldre, Tanel Kaart. Müokardiinfarktiregistrist ja selle andmete kasutamise võimalustest

Krista Fischer. Eluea ja suremuse geneetika ja epidemioloogia: kas pikk eluiga on saatus või valik?

Ants Kaasik. Kodukootud fülogeneetiline meta-analüüsi mudel

Urmas Kõljalg. Elurikkuse andmed ja andmehaldus

Statistiku rollist ravimiuuringutes

Eesti Statistikaseltsi 29. konverentsil 23.-24.03.2018 peetud ettekande põhjal

Jaak Sõnajalg

StatFinn & EPID Research, IQVIA Company

Järgnevate ridadega annan lühikese ja subjektiivse sissevaate ravimiuuringute maailma, kus viimased poolteist aastat statistikuna töötanud olen. Statistikaseltsi konverentsile, ja sealtkaudu ka nende kaante vahele, sattusin kliiniliste uuringutega tegeleva ettevõtte StatFinn & EPID Researchi töötajana. Sellest hoolimata kavatsen hoiduda turunduslikust toonist ja võtan enda eesmärgiks jagada teiste elualade statistikutega oma senise staaži jooksul kogetut.

Pärast mitmeid valusaid õppetunde [1,2] on ravimi katseklaasist turule toomise protsess saanud väga selge struktuuri. Laboris ja loomkatsetes lubavaid tulemusi näidanud ravimi edasiseks uurimiseks võidakse korraldada nn I faasi uuring, kus mõnekümnele tervele vabatahtlikule ravimit manustades uuritakse selle kahjulikke kõrvalmõjusid ja otsitakse inimesele sobivat doosi. I faasi uuringud edukalt läbinud ravimi jaoks saab teha kuni sadat patsienti kaasava nn II faasi uuringu, kus ravimile vastavat näidustust omavatele patsientidele ravimit või platseebot manustades ravimi tõhusust uuritakse. Ravimi turule pääsemise eelduseks on head tulemused nn III faasi uuringus, kuhu kaasatakse juba sadu inimesi ja mille eesmärgiks on veenduda, et uue ravimi massidesse lubamine on õigustatud. Juba turule jõudnud ravimi kasutuskõlblikkuses veendumiseks korraldatakse üha sagedamini nn IV faasi uuringuid, millega soovitakse saada ülevaade kasutusmustritest ja kõrvalmõjude sagedusest kogu populatsiooni seas. Lisaks I-IV faasi uuringutele korraldavad ravimifirmad ka epidemioloogilisi uuringuid, millega soovitakse huvipakkuva näidustuse levimust paremini mõista. Ravimi turule toomise protsess on tootjale väga kulukas (hinnanguliselt 2 miljardit dollarit iga turule jõudnud ravimi kohta [3]) ja enamasti edutu (näiteks on alates 2003. aastast II faasis uuritud üle 200 dementsuse vastast ravimit, millest pole tänaseks ükski edukalt III faasi uuringut läbinud [4]).

Ravimi turukõlblikkuse tõendamine on ravimifirma ülesanne. Ravimi Euroopas turule saamiseks peab ravimi sobivust kinnitama Euroopa Ravimiamet (EMA), Ameerika Ühendriikides on samas rollis sealne Toidu- ja Ravimiamet (FDA). Mainitud ametiasutuste veenmiseks ei piisa reeglina ühest I, II ja III faasi uuringust; igas faasis võidakse teha kümneid uuringuid üle kogu maailma. Sageli leiavad ravimifirmad, et uuringu mõne aspekti toimetamine on kasulik usaldada mõnele lepingulisele organisatsioonile. Näiteks on minu tööandja rolliks uuringus andmehaldusse ja statistikasse puutuvaga tegelemine.

Ravimiuuringu edukuse mõõdikuks on sageli mõni numbriline või kategooriline tulemusnäitaja (ingl *endpoint*, siin ja edaspidi on oskusterminite eestindamisel kasutatud 2012. a magistritööna loodud sõnastikku [5]). Tulemusnäitajad ja edukuse kriteeriumid peavad olema fikseeritud juba enne uuringu toimumist. Kõigile eelnevalt välja valitud kriteeriumitele vastavad tulemused tuleb pärast uuringut ka avaldada. Nii välditakse olukorda, kus uuringu õnnestumisest huvitatud olev uurija esitleb vaid õnnestumisele viitavaid tulemusi ja jätab ebaõnnestumised mainimata.

Statistiku ülesandeks on edukuskriteeriumite hüpoteesidena vormistamine ja hüpoteeside kontrolliks sobivate tööriistade valimine. Uuringu jooksul vaid korra või kaks mõõdetava tulemusnäitaja jaoks võib ravimi ja platseebo võrdlemiseks piisata t-testist (või Wilcoxonit testist); kategoorilise tulemusnäitaja korral Fisherit täpsest testist. Kui ravimigruppe on üle kahe ja kui soovitakse arvesse võtta mitut faktortunnust, võib sobivaks meetodiks olla dispersioonianalüüs (ANOVA) või kovariatsioonanalüüs (ANCOVA); tulemusnäitaja tüüp võib osutada üldistatud lineaarse mudeli (GLM) sobivusele. Paljudes II ja III faasi uuringutes põhinevad tulemusnäitajad korduvatel mõõtmistel, mistõttu osutuvad kasulikeks tööriistadeks segamudel (MM), üldistatud lineaarne segamudel (GLMM) ja üldistatud hinnanguvõrrandid (GEE). Vähiravimite uuringuil põhinevais publikatsioonides püüavad esimesena tähelepanu Kaplan-Meieri elulemuskõverad, samuti on laialt kasutusel Coxi elukestusmudel. Mõne hingetõmbega oleme viidanud pea kõigile Tartu Ülikooli statistika õppekava bakalaureuse- ja magistriõppekavas tutvustatavatele nn traditsioonilistele mudeldamismeetoditele. Ajastu vaimu kohaselt tõstavad kliinilistes uuringute juures üha enam pead ka vähem traditsioonilised, simulatsioonidele ja masinõppele tuginevad meetodid.

Lisaks edukuskriteeriumide statistikakeelde tõlkimisele jääb statistiku vastutusalasse veel palju muud. Näiteks on uuringu õnnestumise seisukohalt ülioluline sobiva valimimahu arvutamine – ka väga hea ravim võib edukuskriteeriumi suhtes põruda, kui uuringusse kaasatud patsientide arv on liiga väike. Veel on statistiku ülesandeks sobiva randomiseerimismeetodi tagamine. Kui korrektne juhustikustamine on mingil

põhjusel võimatu (näiteks retrospektiivsetes uuringutes), võivad vajalikuks osutuda meetodid, kus igale ravimit tarbinud isikule võimalikult sarnane ravimit mitte tarbinud isik vastavusse seatakse. Teatavatel juhtudel on kasutusel imputeerimismeetodid. Pea igas uuringus peab statistik silmas pidama I ja II tüüpi vea suurust ja mitmesest võrdlemisest tulenevat. Kõigele eelnevale lisaks võib uuringu ülesehitusest või uuritavast näidustusest lähtuvalt vajalikuks osutuda mõni hoopis spetsiifiline meetod.

Nagu tulemusnäitajate puhul, tuleb ka kogu statistika-alane plaan juba enne uuringut algust paika panna. See tähendab statistikule põhjalikku planeerimistööd. Enne andmete kogumist tuleb otsustada, milliseid statistilisi meetodeid kogutud andmetele rakendatakse. Sageli tähendab see muuhulgas ka programmikoodide kirjapanemist. Detailselt tuleb planeerida ka kirjeldav analüüs: statistik peab otsustama, millised tabelid ja joonised uuringu lõpuks sünnivad. Veel enam, iga tulevase tabeli jaoks loob statistik tärnide või x-tähtedega täidetud tabelitooriku. Niivõrd detailse planeerimisprotsessi esmane eesmärk on uuringu aususe tagamine, lisaks aitab detailne plaan ka analüüsikoode kirjutavaid ja rakendavaid programmeerijaid.

Statistiku töös on programmeerimine tähtsal kohal. Programmeerimiskeeltest on I-III faasi uuringuis olulisim SAS, IV faasi ja epidemioloogilistes uuringutes on laiemalt kasutusel R. Tulemuste esitlemisel on väga kasulik TeX-dokumentide vormistamise oskus.

Ehkki ravimiuuringutes statistikuna töötades on võimalik oma ülesannetega hakkama saada n-ö andmete taga olevasse - haigustesse, ravimi töömehhanismidesse jms – sisse elamata, on pikemas plaanis usutavasti edukam statistik, kes ka sellel rindel toimuvat mõista üritab. Uudishimulikele ja meditsiinist huvitatutele pakub see töö palju, sh pidevat tutvumist erinevate näidustustega ja uute ravimitega. IV faasi uuringutes ja epidemioloogistes uuringutes pole haruldane, et statistikul on kande roll ka uuringute tulemuste publikatsiooniks kirjutamisel. Selle töö juures on teatavad meditsiinalased teadmised hädavajalikud.

Austraalia statistik Terry Speed on väitnud [6], et statistika ja matemaatika suhtuvad üksteisesse kui kokakunst ja keemia. Nii statistikas kui kokakunstis on lisaks teooriale olulised kogemused ja konteksti tundmine. Ravimiuuringutes leiavad rakendust väga mitmed statistilised meetodid, mistõttu on see valdkond statistikaalaste kogemuste omandamiseks hästi sobiv. Lisaks kasvatab pidev töö meditsiiniandmetega igapäevaselt teadmisi arstiteaduse vallas. Seega võib ravimiuuringute statistiku töö oma mitmekülguses ja vastutusrikkuses statistikule vägagi innustav olla.

Viited

1. Fintel B, Samaras A, Carias, E. The Thalidomide Tragedy: Lessons for Drug Safety and Regulation | Helix Magazine. [tsiteeritud 20. mai 2018]; Kättesaadav aadressil: <https://helix.northwestern.edu/article/thalidomide-tragedy-lessons-drug-safety-and-regulation>
2. Roxby P. Has drug trial safety improved? BBC News [Internet]. 24. mai 2013 [tsiteeritud 20. mai 2018]; Kättesaadav aadressil: <http://www.bbc.com/news/health-22556736>
3. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ. mai 2016;47:20–33.
4. FDA wants to help unproductive drugmakers. The Economist [Internet]. 22. märts 2018 [tsiteeritud 24. märts 2018]; Kättesaadav aadressil: <https://www.economist.com/news/business/21739179-americas-pharmaceutical-giants-may-also-need-rethink-their-business-model-fda-wants-help>
5. B. P. Inglise-eesti kliiniliste uuringute sõnastik [Internet] [magistritöö]. Tartu Ülikool; 2012 [tsiteeritud 20. mai 2018]. Kättesaadav aadressil: http://dspace.ut.ee/bitstream/handle/10062/27441/Prakson_Birgit_2012.pdf
6. Speed T. Questions, Answers and Statistics. International Conference on Teaching Statistics II [Internet]. Victoria, Kanada: IASE; 1986. Kättesaadav aadressil: https://iase-web.org/Conference_Proceedings.php?p=ICOTS_2_1986

Sündimusest, sündimuse mõõtmisest ja sündimuse mudelitest

Eesti Statistikaseltsi 29. konverentsil 23.-24.03.2018 peetud ettekande põhjal

Ene-Margit Tiit

Tartu Ülikooli emeriitprofessor, Eesti Statistikaamet

Sündimus on rahvastikuprotsess, mis Eesti üldsusele alati väga suurt huvi pakub. Tuleb aga tõdeda, et kõik rahvastikuprotsessid on omavahel seotud ja sündimuse käsitlemisel on oluline jälgida ka teiei protsesse. Huvitav on aga see, et erinevatel ajaperioodidel on sündimuse sõltuvus muudest protsessidest olnud erinev ja igasuguste sündimuse mudelite loomisel on oluline jälgida ajastuomaseid suundumusi.

Väga pika aja jooksul – hinnanguliselt 200 000 aasta jooksul alates *homo sapiensi* tekkest olid sündimus ja suremus praktiliselt tasakaalus. Inimese elukestus oli sellel 20—30 aastat ja inimkond kasvas iga põlvkonnaga keskmiselt 0,07%. Kindlasti oli rahvastikurühmi, kes kasvasid kiiremini, kuid oli ka hääbuvaidning välja surevaid rühmitusi, hoolimata sellest, et sündimus (samuti nagu suremus) oli toona tänapäevasega võrreldes väga kõrge.

Äkiline muutus toimus aga mõnesaja aasta eest, 18. sajandil. Ühelt poolt arenes tootmine sedavõrd, et inimesi enam ei ähvardanud pidev näljasurma oht. Teiselt poolt hakkas tekkima esmane arusaamine tervishoiust ja hügieenist, tekkis harjumus pesta käsi. Sellel oli vapustav tagajärg – suremus langes, kuid esialgu jäi sündimus sama kõrgele tasemele, nagu varem. Rahvaarv hakkas kiiresti kasvama – kõigepealt kõige arenenumas piirkonnas – Lääne-Euroopas, siis aga tasapisi mujalgi. Sündimus püsis kõrgtasemel paari põlvkonna vältel ja langes siis enneolematult madalale tasemele. Kuna see protsess, mida hakati nimetama demograafiliseks üleminekuks, haaras aina uusi ja uusi rahvaid, riike ja piirkondi, jätkus Maa rahvastiku aina kiirenev kasv. Kasvukiirus saavutas haripunkti möödunud sajandi 60ndatel aastatel, mil rahvaarv suurenes enam kui 2% võrra aastas. Selle järel hakkas kasvukiirus alanema, kasv siiski jätkus.

Rahvastikuteadlasi on alati huvitanud rahvaarvu ennustamine – seda ootavad neilt ka riigijuhid ja ühiskonnategelased. Tuleb aga tõdeda, et rahvaarvu ennustamine on väga tänamatu töö. Kõigist ajaloo vältel tehtud rahvastikuproгноosidest on täppi läinud nväga vähesed, enamik on mööda läinud, sealjuures tihti väga rängalt. Seda tõsiasja illustreerivad ÜRO rahvastikuekspertide poolt käesoleva sajandi lõpuks tehtud ennustused, mis üksteisest radikaalselt erinevad. Madal prognoos näitab rahvaarvu kahanemist ja langust alla 7 miljardi taseme; keskmine ennustab mõõdukat kasvu veidi üle 10 miljardijoone; kõrge aga ennustab rahvaarvu suurenemist 16 miljardini. Seega on kõrge ja madala prognoosi vahe enam kui kahekordne, kuid kõige tõenäolisem on siiski vaheline areng, mis vastab keskmisele arengustenaariumile.

Suremuse kõige ilmekam indikaator on oodatav eluiga sünnil (elukestus). See on näitaja, mis on seotud konkreetse aastaga ja näitab, kui kaua elaks sel aastal sündinud laps eeldusel, et kogu tema eluaja jooksul jääb suremus sellele tasemele, nagu on vaatlusaastal. Tegelikult pole see eeldus täidetud ja vaadeldaval aastal sündinud lapsed elavad üldjuhul märksa kauem kui näitab selle aasta kohta arvatud oodatav eluiga. Kogu maailmas on viimase paarisaja aasta jooksul oodatav eluiga kahekordistunud ja üldiselt jätkab kasvamist, samal ajal erinevused riikide ja piirkondade vahel vähenevad, vaid Aafrika kesk- ja lõunapoolsetel aladel on oodatav eluiga märksa madalam kui mujal maailmas. Selles piirkonnas püsib ka sündimus kõrgtasemel ja rahvastikuareng on võrreldav Euroopaga kolmsada aastat tagasi. Kuigi olukord pole minevikuga päriselt võrreldav – kõrget suremust Aafrikas põhjustab kõige enam AIDS (hoolimata tänapäevase meditsiini mõningasest mõjust).

Mis mõjustab oodatavat eluiga? Võrdlemisi tugev (mittelineaarne) seos on oodataval elueal riigi jõukusega (SKP isiku kohta) – mida kõrgem on riigi jõukuse tase, seda pikem selle elanike keskmine eluiga. See seos on igati ootuspärane, sest on selge, et jõukamas riigis on tervishoiutingimused paremad ja ka nälja ning alatoitluse oht väike.

Sündimuse ennustamiseks on aga tarvis kokku leppida, missugust näitajat kasutada sündimuse mõõtmiseks. Igati loomulik on kasutada naise poolt elu jooksul sünnitatud laste arvu – see on nn lõpetatud sündimuskordaja – kuid tegelikus elus seda siiski eriti palju ei kasutata. Selle näitaja puuduseks on ajastatusehiline, sest selle näitaja arvutamine on võimalik sünnitamiseast juba väljunud naiste puhul. Praktiliselt kasutatakse sünnitamisaktiivsuse hindamisel summaarset sündimuskordajat, mis mõõdab naise poolt sünnitatavat laste arvu eeldusel, et sünnitamisaktiivsus jääb kogu naise viljaka elu jooksul samaks. Kui see eeldus on täidetud, siis langevad mõlemad nimetatud mõõdikud täpselt kokku. Tegelikuses pole see eeldus tavaliselt täidetud ja selletõttu sõltub summaarne sündimuskordaja oluliselt sünnituse muutustest. Neil perioodidel, kui keskmine sünnitamisvanus väheneb ja sünnitatakse järjest nooremas eas, kasvab summaarne sündimuskordaja ja selle väärtus näitab naise kohta keskmiselt suuremat sünnitatud laste arvu kui seda tagantjärele kinnitab lõpetatud sündimuskordaja. Samal ajal aga rahvastiku arvukus suureneb, sest põlvkondade vahed vähenevad ja üheaegselt on elus rohkemate põlvkondade esindajad. Kui aga keskmine sünnitamisvanus suureneb ja sündide lükatakse edasi, kahaneb summaarne sündimuskordaja ja väheneb ka rahvaarv hoolimata sellest, et lõpetatud sündimuskordaja ei tarvitsegi väheneda. Põhjuseks on sel juhul põlvkondade vahede suurenemine.

Kui vaadata summaarse sündimuskordaja väärtusi kogu maailmas, siis on näha, et selle väärtus on suur nimelt nendes maailma piirkondades, kus oodatav eluiga on lühike – peamiselt Aafrikas. See kinnitab taas, et seal on rahvastikuareng demograafilise ülemineku eelsel tasemel. Siiski näitab statistika, et summaarne sündimuskordaja alaneb kõikjal – välja arvatud arenenud riigid, kus see on enam-vähem stabiliseerunud – ja läheneb ühisele väärtusele, mis on enam-vähem võrdne taastetasemega, so pisut üle arvu 2.

Summaarse sündimuskordaja väärtus on võrdlemisi hästi ennustatav inimarengu indeksi HDI alusel. See indeks, mille üks komponent on riigi jõukus (SKP isiku kohta), kuid sellele lisanduvad veel oodatav eluiga ja haridustase, omandab väärtusi 0 ja 1 vahel ja jagab kõik maailma riigid nelja kategooriasse: väga kõrge, kõrge, keskmise ja madala inimarengu tasemega riigid. Inimarengu indeks arvutatakse ÜRO poolt kõigi riikide jaoks igal aastal ja riigid järjestatakse selle järgi. Selgub, et Mida kõrgem on inimarengu indeksi väärtus, seda madalam on keskmine summaarne sündimuskordaja selles riigis. Seda üldist seaduspära siiski rikub käesoleva sajandi alguses ilmnenud nähtus, mida kirjeldab ÜRO juures töötava Soome päritolu rahvastikustatistiku nn J-kõver: väga kõrge inimarengu indeksiga riikides summaarne sündimuskordaja üldjuhul ei alane inimarengu indeksi kasvades, vaid pigem hakkab tõusma. Tulemus on sisulises mõttes ootuspärane – ei saa ju loogiliselt olla tõene olukord, et ülikõrge inimarengu tasemega riikide rahvastik lihtsalt demograafilistel põhjustel välja sureb. Summaarse sündimuskordaja kahanemise pidurdumise põhjusena tuleb arvesse ka see, et niisugustes riikides on sündide edasilükkamine peatunud, sest keskmine sünnitamisvanus on tõusnud juba mõistliku loomuliku piiri lähedale. Igal juhul näitab lihtne regressioonanalüüs, et riikides, kus inimarengu indeksi väärtus on vähemalt 0,85, on korrelatsioon summaarse sündimuskordaja ja inimarengu indeksi vahel positiivne ja indeksi kasvades toimub ka sündimuskordaja pisike tõus. Tuleb aga tõdeda, et väga kõrge inimarengu indeksiga riigid (nendeseas ka Eesti) on igas suhtes üsna erinäolised, olulisel määral mõjutab sündimuskäitumist kultuuritaust, sh religioon. Eriti kõrge on sündimus Iisraelis, väga madal aga ida-Aasia riikides – Lõuna-Koreas, Jaapanis, Singapuris ja Hong-Kongis. Euroopa riikidest kuuluvad sellesse rühma lisaks katoliiklikule Irimaale ja traditsiooniliselt kõrge sündimusega Islandile ka teised Põhjamaad. See on lohutav ja kainestav tõik.

Missugused on aga rahvastikunäitajad, millega tänapäeva sündimusnäitajad on seotud? Üsna üllatuslikult on keskmine sünnitamisvanus kogu maailmas võrdlemisi ühtlane, olles ümmarguselt 30 aasta juures. Tõsi, kõrgemate sündimusnäitajatega maades alustatakse sünnitamist nooremas eas, kuid sünnitamist jätkatakse ka peale kolmekümnendaid aastaid. Madalama sünnitamisaktiivsusega riikides sünnitatakse üks või kaks last nimelt kolmekümneaastaselt.

Väga oluliselt erineb aga riigiti mediaanvanus – see on niisugune vanus, millest pooled riigi elanikud on nooremad ja pooled vanemad. Riikides, kus on palju lapsi ja noori, on mediaanvanus madal. Seevastu riikides, kus noorte oskaal on suhteliselt madal ja elanikkond võrdlemisi eakas, on ka mediaanvanus kõrge. Praegusel ajal varieerub mediaanvanus riikide lõikes 15st kuni 45 aastani.

Omamoodi kõnekas näitaja on keskmise sünnitamisvanuse ja elanikkonna mediaanvanuse vahe. Kui see on positiivne, st sünnitatakse vanemana kui mediaanvanus, muutub küsitavaks, kui paljud emadest üldse näevad oma laste täiskasvanuks sirgumist. Vastupidi, kui sünnitamisvanuse ja mediaanvanuse vahe on negatiivne, st et sünnitatakse mediaanvanusest nooremana, võib ema ka pärast laste üleskasvatamist vanaduspõlve

nautida. Huvitaval kombel on see keskmise sünnitamisvanuse ja elanikkonna mediaanvanuse vahe näitaja, mis on tugevasti korreleeritud summaarse sündimuskordajaga, $r = 0,894$. Seega kirjeldab alljärgnev lineaarne mudel 80% summaarse sündimuskordaja hajuvusest riigiti:

$$\text{TFR} = 0,147 \times (\text{keskmise sünnitamisvanus} - \text{elanikkonna mediaanvanus}) + 2,85.$$

Kokkuvõttes tuleb tõdeda, et demograafiline käitumine ei ole kõigil ajaloo perioodidel olnud ühesugune. See tõsiasi teeb keerukaks niihästi rahvastikuprognoside tegemise kui ka perepoliitika kavandamise. Demograafilist käitumist mõjutavad niihästi bioloogilised kui ka keskkonnategurid, kuid samuti ka väärtushinnangud ja käitumisnormid. Perekonnaloomealiste protsesside areng toimub omasoodu ja endiste aegade olukordade ning tõekspidamiste juurde tagasi pöördumine ei õnnestu.

Viitekirjandus

1. Myrskylä, Mikko and Kohler, Hans-Peter and Billari, Francesco C. (2009) Advances in development reverse fertility declines *Nature*, 460 (7256). 741-743. ISSN 0028-0836;
2. Myrskylä, Mikko and Kohler, Hans-Peter and Billari, Francesco C. (2011) High development and fertility: fertility at older reproductive ages and gender equality explain the positive link MPIDR Working Papers, Max Planck Institute for Demographic Research, Rostock, Germany.
3. Riikide pingerida inimarengu indeksi järgi
https://et.wikipedia.org/wiki/Riikide_pingerida_inimarengu_indeksi_j%C3%A4rgi.
4. UN. World Population Prospects. <https://esa.un.org/unpd/wpp/>
5. World Population Growth. <https://ourworldindata.org/world-population-growth>
6. Demographic transition. https://en.wikipedia.org/wiki/Demographic_transition
7. Projections of population growth.
https://en.wikipedia.org/wiki/Projections_of_population_growth
8. Life expectancy. https://en.wikipedia.org/wiki/Life_expectancy
9. Max Roser. Life Expectancy . <https://ourworldindata.org/life-expectancy>.
10. Life expectancy versus GDP per capita. <http://www.statisticalconsultants.co.nz/blog/life-expectancy-at-birth-versus-gdp-per-capita-ppp.html>

Müokardiinfarktiregistrist ja selle andmete kasutamise võimalustest

Eesti Statistikaltsi 29. konverentsil 23.-24.03.2018 peetud ettekande põhjal

Gudrun Veldre¹, Tanel Kaart²

¹ SA Tartu Ülikooli Kliinikum, Müokardiinfarktiregister; Tartu Ülikool, kliinilise meditsiini instituut

² Eesti Maaülikool, veterinaarmeditsiini ja loomakasvatuse instituut

Sissejuhatus

Südameveresoonkonna haigused on Eestis, nagu Euroopa Liidus ja kogu maailmas, peamiseks surmade põhjuseks. Seejuures aga esineb Eestis südameveresoonkonna haiguseid ligi kaks korda sagedamini kui Euroopas keskmiselt [1].

Südameveresoonkonna haigustest kõige sagedasemaks on südame isheemiatõbi, mis seisneb enamasti ühe või mitme südame pärgarteri (koronaararteri) ahenemises või sulgumises. Enamasti on see ahenemine või sulgumine tingitud aterosklerootilistest muutustest koronaararterites ja põhjustab südamelihase hapnikuvarustuse vähenemist (isheemiat). Kriitiline ja ulatuslik verevarustuse vähenemine võib põhjustada südamelihase (müokardi) infarkti või koguni surma [2]. Infarkt on kudede kärbus e. nekroos, mis tekib südamelihase hapniku puuduse tagajärjel minutite ja tundidega. Hävinud südamelihase asemele tekib südame pumbafunktsioonis mitteesalev armkude.

Aeg on infarkti ravi alustamiseks kriitilise tähtsusega, et patsiendi südamelihase ei häviks ja ei asenduks armkoega. Seetõttu peab ägeda müokardiinfarktiga patsient jõudma nii kiiresti kui võimalik haiglasse, kus on kaasaegse invasiivse ravi võimalused. Eestis on kaasaegse invasiivse ravi – perkutaanse koronaarintervensiooni (PKI) – 24/7 võimalusega haiglaid kolm: Tallinnas Põhja-Eesti Regionaalhaigla ja Ida-Tallinna Keskhaigla ning Tartus Tartu Ülikooli Kliinikum. Neile kolmele haiglale lisandub kaks haiglat (Pärnu Haigla ja Ida-Virumaa Keskhaigla), kus on PKI võimalus kord nädalas.

1. Müokardiinfarktiregister – mis, kes ja milleks

Müokardiinfarktiregister (MIR) on Eesti Vabariigi valitsuse määrusega alates 1. jaanuarist 2012 ellu kutsutud riiklik andmekogu, mille vastutav töötleja on Sotsiaalministeerium ja volitatud töötleja on Tartu Ülikooli Kliinikum, kus on müokardiinfarktiregistri pidamiseks loodud spetsiaalne teenistus Müokardiinfarktiregister.

Müokardiinfarktiregister riikliku andekoguna sai alguse 2012. aastal, kuid eeltööd registri loomiseks alustati juba 2000. aastal. Andmeid on registris 2006. aastast, mil register töötas regionaalhaiglate kokkuleppel eraõigusliku registrina.

MIR-i peetrake eelkõige selleks, et

- 1) parandada müokardiinfarkti diagnostikat ja ravikvaliteeti Eestis;
- 2) aidata kaasa ägeda müokardiinfarkti (ÄMI) patsientide suremuse vähendamisele;
- 3) hõlbustada tervishoiu teenuste planeerimist;
- 4) hõlbustada müokardiinfarktijuhtude statistika korraldamist;
- 5) viia läbi epidemioloogilist uurimistööd ja teadusuuringuid.

Statsionaarselt ravitud ÄMI patsientide teatiste sisestamine registrisse on kohustuslik kõikidele tervishoiuteenuste osutajatele, kes ravivad Eestis ÄMI patsiente, s.o. patsiente Rahvusvahelise Haiguste diagnoosikoodidega (RHK-10 diagnoosikoodiga) I21 ja I22.

Praegu on Eestis kokku 20 haiglat, kes osutavad ägeda müokardiinfarkti patsientidele aktiivravi. Nendest 20 haiglast (täpsemalt vt. registri kodulehelt www.infarkt.ee) on registril kokku 260 MIR infosüsteemi kasutajat (esitavad ja vastutavad kasutajad).

Register on unikaalne ka selle poolest, et alates 26.11.2012 on registril Teadusnõukogu, mis nõustab MIR vastutavat töötajat ja hoiab silma peal registri heal käekäigul.

2. Registre kogutavate andmete koosseis

Registre kogutakse andmeid, mis on tavalised ÄMI patsientide haigusloos, ning andmekoosseis on vastavuses Euroopa Kardioloogide Seltsi eestvedamisel Euroopa Liidu tasandil 2004. a. vastuvõetud konsensusdokumendiga CARDS (*Cardiology Audit and Registration Data Standards*) [3].

Registre kogutakse järgmiseid andmeid:

- isikuandmed;
- pikkus ja kaal
- kardiovaskulaarsed riskifaktorid, varasem anamnees;
- andmed patsiendi kliinilise seisundi kohta haiglasse saabumisel;
- EKG kirjeldus;
- ataki algus (kuupäev, kellaaeg);
- haiglas kasutatud ravimid;
- uuringud: ehhoKG, koronarograafia;

- revaskulariseerimine: aorto-koronaarne šunteerimine (AKŠ) või perkutaanne koronaarinterventsioon (PKI);
- labori analüüside tulemused (kolesteroolid, triglütseriidid);
- haiglasised tüsistused;
- soovitud ambulatoorseks raviks.

Samuti sisestatakse registrisse lõplik ÄMI diagnoos [4] ning surnud patsientidel lisanduvad lahangu andmed. Kokku on MIR-s ühe patsiendi kohta registreeritud kuni 90 näitajat.

3. Lühiülevaade registrisse kogutud andmetest

Kokku on aastail 2012-2016 registrisse sisestatud üle 12 000 teatise (Tabel 1), sellele lisanduvad ca 3000 teatist eelmisest, 2017. aastast, ja alates 2006. aastast olevad andmed. Riiklikule registrile eelnevast perioodist on registris 11 543 kinnitatud teatist.

Tabel 1. Aastail 2012-2016 müokardiinfarktiregistrisse sisestatud ÄMI teatised ja juhud.

Aasta	Teatiste arv	Juhtude arv	TTO-de arv
2012	3263	2927	18
2013	2939	2593	19
2014	2969	2598	18
2015	3210	2827	18
2016	3150	2780	18

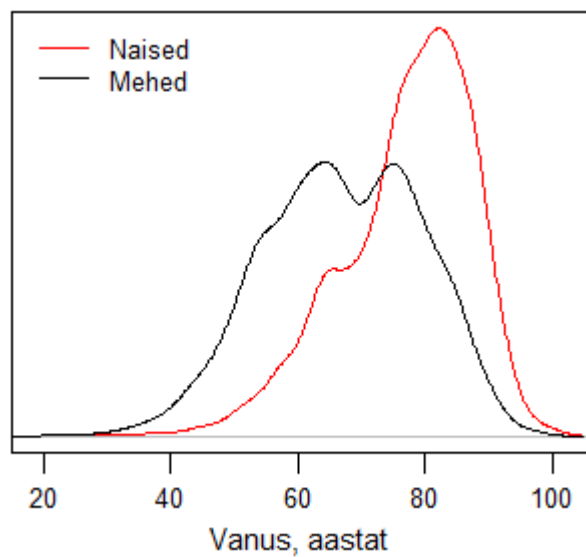
TTO – tervishoiuteenuste osutaja

Keskmine ÄMI patsient on 71 aastane, meeste ja naiste keskmise vanuse erinevus on ca 10 aastat (Tabel 2). Vanuse varieerumine on aga üsna suur. Kõige noorem patsient on 19-aastane, kõige vanem 102-aastane. Seejuures on ÄMI patsientide vanusejaotusel (Joonis 1) bimodaalne struktuur, mis on eriti selge meeste puhul – ÄMI risk on kõrgeim pisut peale 60. eluaastat, misjärel langeb ja saavutab uue tipu pisut enne 80. eluaastat.

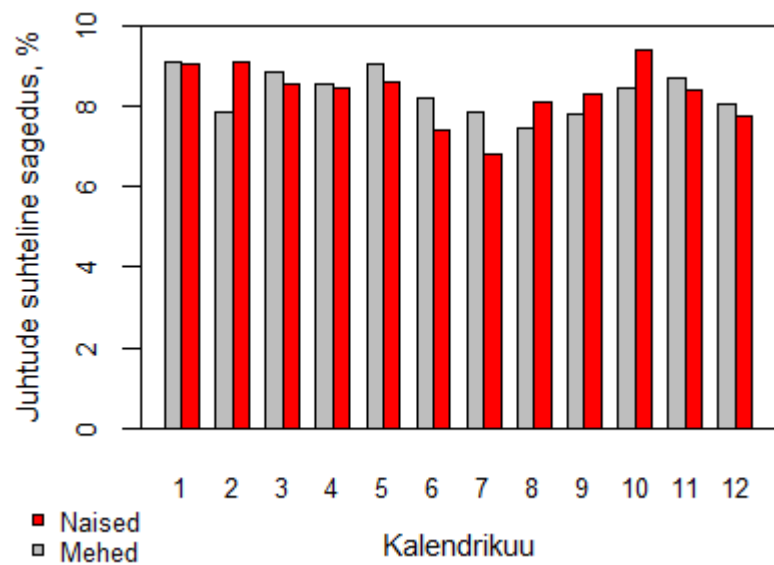
ÄMI juhtude aastasiseselt jaotuselt (Joonis 2) ilmneb, et enam juhte esineb kevadel ja sügisel ja vähem suvel. Nädalapäevade osas oli ÄMI juhtude sagedus suurim teisipäeviti ja kellaaegade osas ajavahemikus 10-13.

Tabel 2. Aastail 2012-2016 registresse kantud ÄMI patsientide vanus (aastates).

Aasta	Keskmine	Mehed	Naised
2012	71,8	67,6 (19–99)	77,6 (29–100)
2013	70,8	66,5 (22–96)	76,8 (22–101)
2014	71,2	67,2 (22–99)	76,9 (35–100)
2015	71,7	67,4 (25–98)	77,5 (36–102)
2016	71,3	67,2 (23–97)	77,0 (28–100)

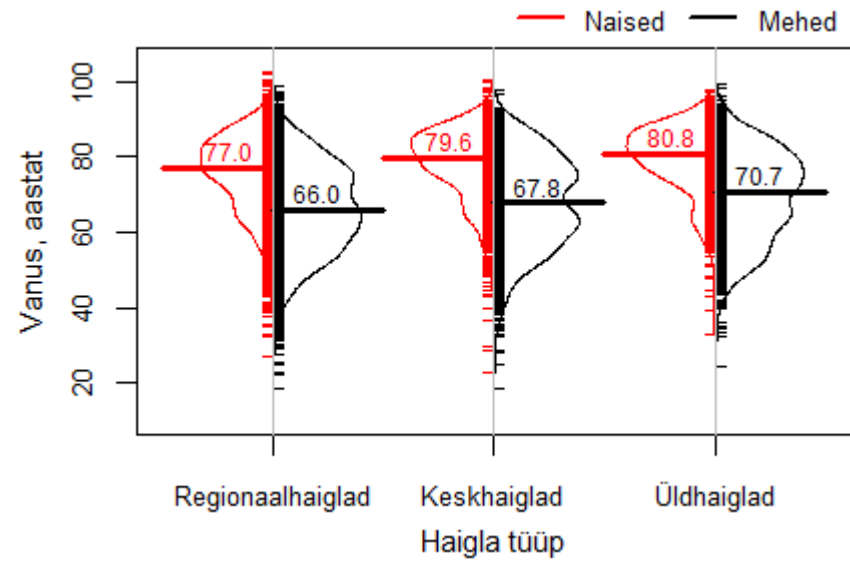


Joonis 1. ÄMI patsientide vanusejaotus sugude kaupa.

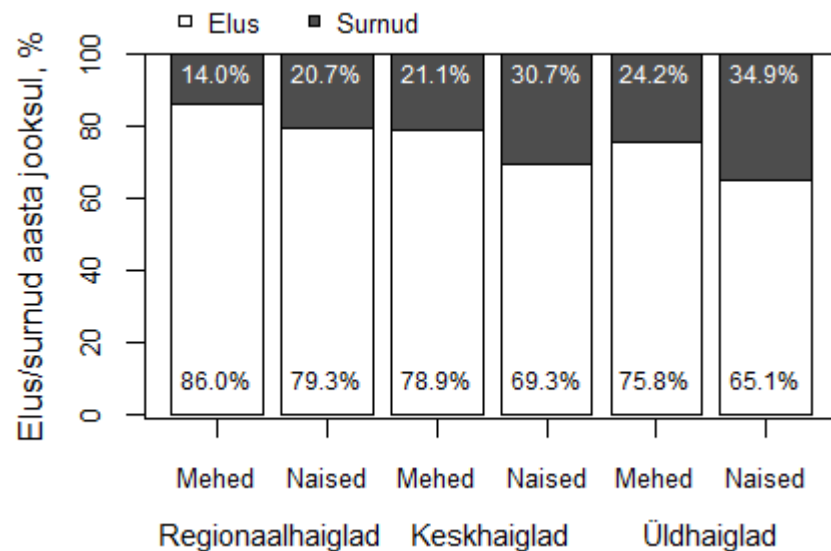


Joonis 2. ÄMI juhtude aastasine jaotus.

Haiglatüüpide osas on näha selget trendi – valdavalt väiksemates kohtades paiknevate üldhaiglate ÄMI patsiendid on keskmiselt vanemad, samuti on seal suhteliselt vähem noori patsiente, seda eriti naiste puhul (Joonis 3). Samuti on raskes seisundis haiglasse sattunud patsiente üldhaiglates rohkem, mis kokkuvõttes väljendub ka patsientide elulemuses (Jooni 4). Joonisel 4 on näha ka, et kuigi kõigist ÄMI patsientidest on 57,8% mehed, on suurem kõrgem hoopis naiste seas.



Joonis 3. ÄMI patsientide vanusejaotus sugude ja haiglatüüpide kaupa; paksu horisontaalse joonega ja arvuliselt on esitatud vanuse mediaan.



Joonis 4. ÄMI patsientide elulemus ühe aasta jooksul haiglatüüpide ja sugude kaupa.

4. MIR-i andmete kasutamine

Hoolimata oma suhteliselt lühikesest ajaloost on MIR-i andmed lisaks iga-aastastele aruannetele (vt <http://www.infarkt.ee/et/aruanded>) leidnud kasutust ka mitmetes ettekannetes ja teadustöodes.

Eesti Kardioloogide Selts ja selle töörühmad on aktiivselt kasutanud MIR-i andmeid ettekanneteks arstidele ning oma töö paremaks organiseerimiseks. MIR-i andmetel on tehtud ettekandeid mitmel rahvusvahelisel kardioloogilistel konverentsidel.

Registri pilootprojekti järgselt 2005. aastal kaitses Tiia Ainla oma doktoritöö „Äge müokardiinfarkt Eestis: kliinilised tunnused, ravikäsitlus ja -tulemused“. Selle töö raames kogutud andmed ja töö tulemused panid suuresti aluse MIR-i tekkele. Aastal 2013 kaitses nii registri eraõigusliku perioodi andmetel kui ka registri, haigekassa ja rahvastikuregistri andmebaasi linkimisel saadud andmetel põhineva doktoritöö „Äge müokardiinfarkt Eestis: muutused kliinilistes tunnustes, ravikäsitluses ja -tulemustes“ Mai Blöndal. Hetkel on valmimas Aet Saare doktoritöö „Äge müokardiinfarkt Eestis - käsitlus ja ravitulemused viimase kümnendi jooksul“.

Haiglaeelseid ja -siseseid viivitusi infarktihaigete ravis uuris Tiina Mark oma 2014. aastal kaitstud magistritöös „STEMI-patsientide haiglaeelse ja -sisese reperfusioonravi viivitused“. Infarkti esinemissageduse ajalist dünaamikat ja seoseid ilmaga uuris oma residentuuri uurimistöös „Ilm kui

müokardi infarkti riskifaktor“ 2015. aastal Jana Alep-Möls. Ajalisi viivitusi uuris 2017. aastal ka Joonas Tuutma, kelle uurimistöö „ST-elevatsiooniga müokardiinfarktiga patsientide ajalised viivitused Tartu Ülikooli Kliinikumis aastal 2016“ (kaasautorid Olga Kiss, Gudrun Veldre ja Jaan Eha) tulemused kanti ette TÜ Arstiteaduskonna 385. aastapäeva puhul korraldatud teaduskonverentsil ning ettekanne ise pälvis auhinna kategoorias „Parim üliõpilaste suuline ettekanne“.

Aastal 2015 kasutas Sven Laur võrdlusetalonina MIR-i andmeid ESS-i 27. Konverentsil TÜ ja Tarkvara Tehnoloogia Arenduskeskus OÜ ettekandes „Digiloo andmete kasutamise võimalikkusest statistilistes uuringutes“ (kaasautorid Dage Särg, Aleksandr Tkatchenko, Timo Petmanson ja Jaak Vilo).

Kokkuvõte

MIR on, võrreldes mitmete teiste Eesti registritega, veel suhteliselt noor andmekogu, aga sisaldab võib-olla just tänu oma noorusele (õppimisele teiste kogemustest) ja eelkõige tänu hästi läbi mõeldud ning kohati ambitsioonikalegi plaanile koguda ja talletada kogu Eestit hõlmavalt südamehaiguste konkreetse, laialt levinud indikaatori diagnoosiga patsientide riskifaktoreid, ravi ja selle tulemuslikkust puudutavat infot. Lisaks on MIR-i andmed lingitavad teiste Eesti registrite ja andmekogudega (surma põhjuste register, Geenivaramu andmebaas jmt). Seeläbi on tegu üsna unikaalse ja laialdasi uuringuid teostada võimaldava registriga. Juba praegu kogunenud andmete alusel on võimalik uurida paljusid ÄMI esinemise ja raviga seotud rohkem ja vähem spetsiifilisi meditsiinilisi aspekte, mille osas loodavad allakirjutanud jätkuvat arstitudengite ja nende juhendajate poolset huvi. Samas on tegu ka väga mitmetasandilise ning omavahel keerukates seostes olevate andmete koguga. Seetõttu ei ole MIR-s sisalduvata andmete analüüs ning nii patsientidele kui ka haiglatele kasulike järelduste tegemine võimalik ilma komplekssete statistiliste analüüsideta. Erinevad MIR-i osad pakuvad reaalseid andmeid nii kitsamateks konkreetsetel statistikameetoditel baseeruvateks analüüsideks kui ka laialdasemateks keerukate mudelite ja mitmemõõtmelise statistika algoritmidega „mängimist“ eeldavateks uuringuteks.

Kokkuvõttes loodame, et antud lühiülevaade MIR-i andmetest ja nende senisest kasutamisest aitab äratada huvi riikliku andmekogu vastu ning ärgitab kasutama MIR-i andmeid matemaatilise statistika eriala tudengitöodes. Mis omakorda loob aluse kiirema ja mitmekülgsema tagasiside andmiseks meedikutele. Register on midagi väärt üksnes siis, kui seda kasutatakse!

Kasutatud kirjandus

[1] https://intra.tai.ee//images/prints/documents/149787347942_Health%20in%20the%20Baltic%20Countries.pdf

[2] Eha, J.; Ruusalepp, A.; Palm, E.; Ringmets, I.; Kiiwet, R-A. (2011). Revaskulariseerimine südame isheemiatõve ravis Eestis. Tartu Ülikool, Tervishoiu toimetised, 1/2011, 6.

[3] https://www.escardio.org/static_file/Escardio/EU-affairs/CARDS-dataset-PCI-1104.pdf

[4] Marandi, T. (töörühma juht) (2013). Müokardiinfarkti kolmas universaalne definitsioon. Eesti Arst, 92(4), 231–235.

Eluea ja suremuse geneetika ja epidemioloogia: kas pikk eluiga on saatus või valik?

Krista Fischer

TÜ Matemaatika ja statistika instituut

TÜ Genoomika instituut, TÜ Eesti Geenivaramu

Eluiga – kõige olulisem ja lõplikum tervisenäitaja?

- Keskmise eluiga on aastate jooksul kasvanud, kuid sellegipoolest näeme märkimisväärselt varieeruvust
- Kui õnnetusjuhtumid välja arvata, siis on keskmisest lühema eluea põhjuseks enamasti mõni haigus
- Erinevate haiguste riskitegurid on erinevad – seega ei saa otseselt ühe või teise riskiteguri põhjal prognoosida eluiga
- Uurides otseselt eluiga (st üldsuremust), saame leida nn universaalseid riski mõjutavaid tegureid, mille kahjulikud mõjud on kokkuvõttes suuremad kui nende võimalik kasutegur

TÜ Eesti Geenivaramu (TÜ EGV) ja teised biopangad – kas neid saab kasutada eluea uurimiseks?

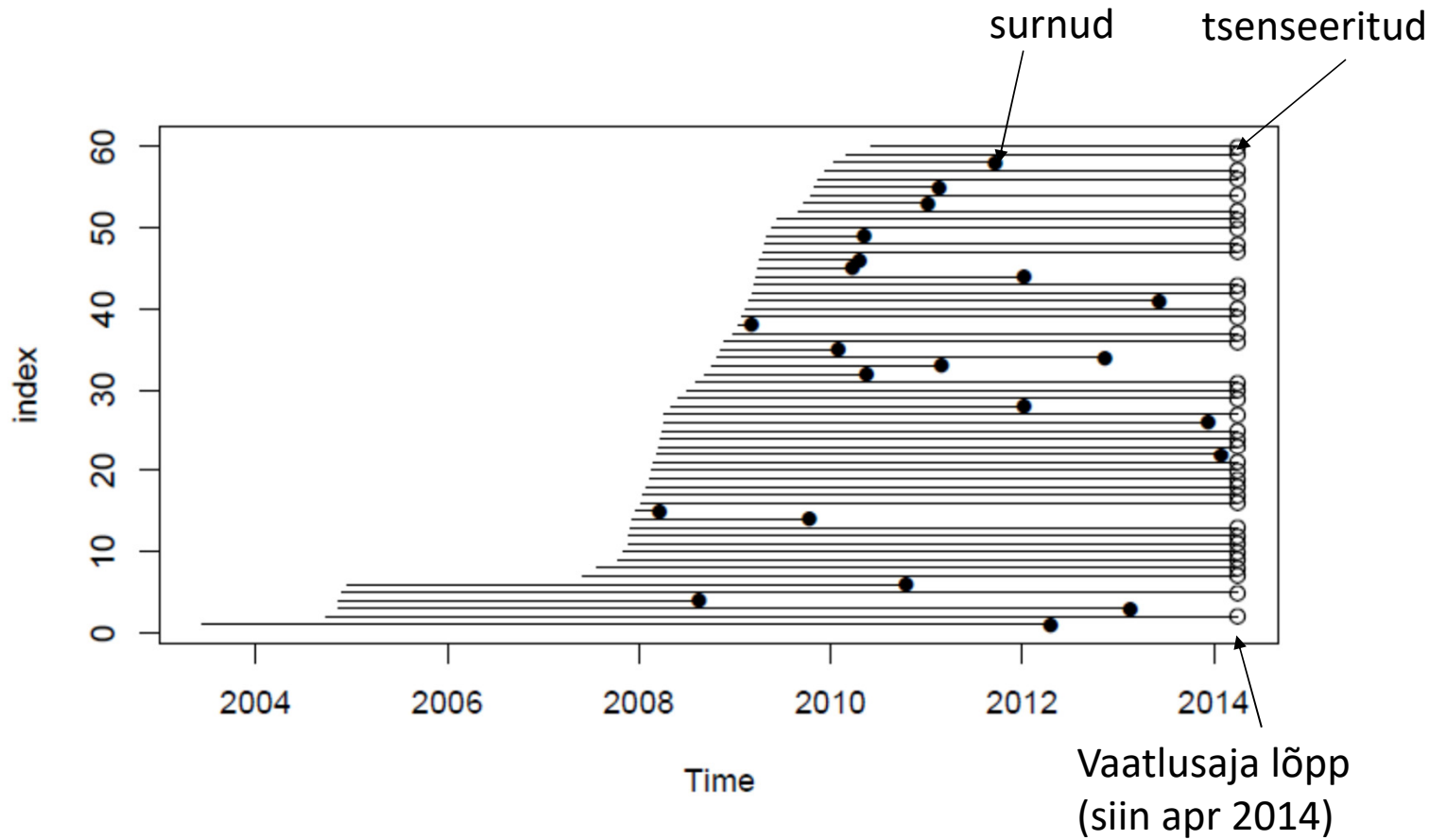
- TÜ EGV: kohordi suurus 53000, 82% inimestest liitunud ajavahemikus 2003-2010, vanus liitumisel 18-103 (keskmise 46)
- UK Biobank (UKBB): kohordi suurus 500000, kaasatud vahemikus 2006-2010, vanuses 40-69

TÜ EGV kohordist on surnud 4284 inimest (juuni 2017), UKBB>8000

Siiski on see suhteliselt väike valimisuurus kestusandmete analüüsiks...

I Ajaskaala valik elukestusanalüüsiks

Näide: 60 indiviidi ja nende jälgimisaeg



Enimkasutatav elukestusanalüüsi meetod: Coxi võrdeliste riskide mudel

Riskifunktsioon (Hazard function):

$$h(t) = \lim_{dt \rightarrow 0} P(t \leq T < t + dt | T > t)$$

Üleelamisfunktsioon (ka: elulemusfunktsioon,
Survival function):

$$S(t) = P(T > t) = e^{-\int_0^t h(u) du}$$

$$h(t|X_1, \dots, X_k) = h_0(t) e^{X_1\beta_1 + \dots + X_k\beta_k}$$

$$= h_0(t) \times e^{X_1\beta_1} \times \dots \times e^{X_k\beta_k}$$

Riski(määrade)suhe (*hazard ratio*) mis vastab ühikulisele erinevusele X_j -s:

$$HR_j = e^{\beta_j}$$

Osalise tõepära funktsioon:

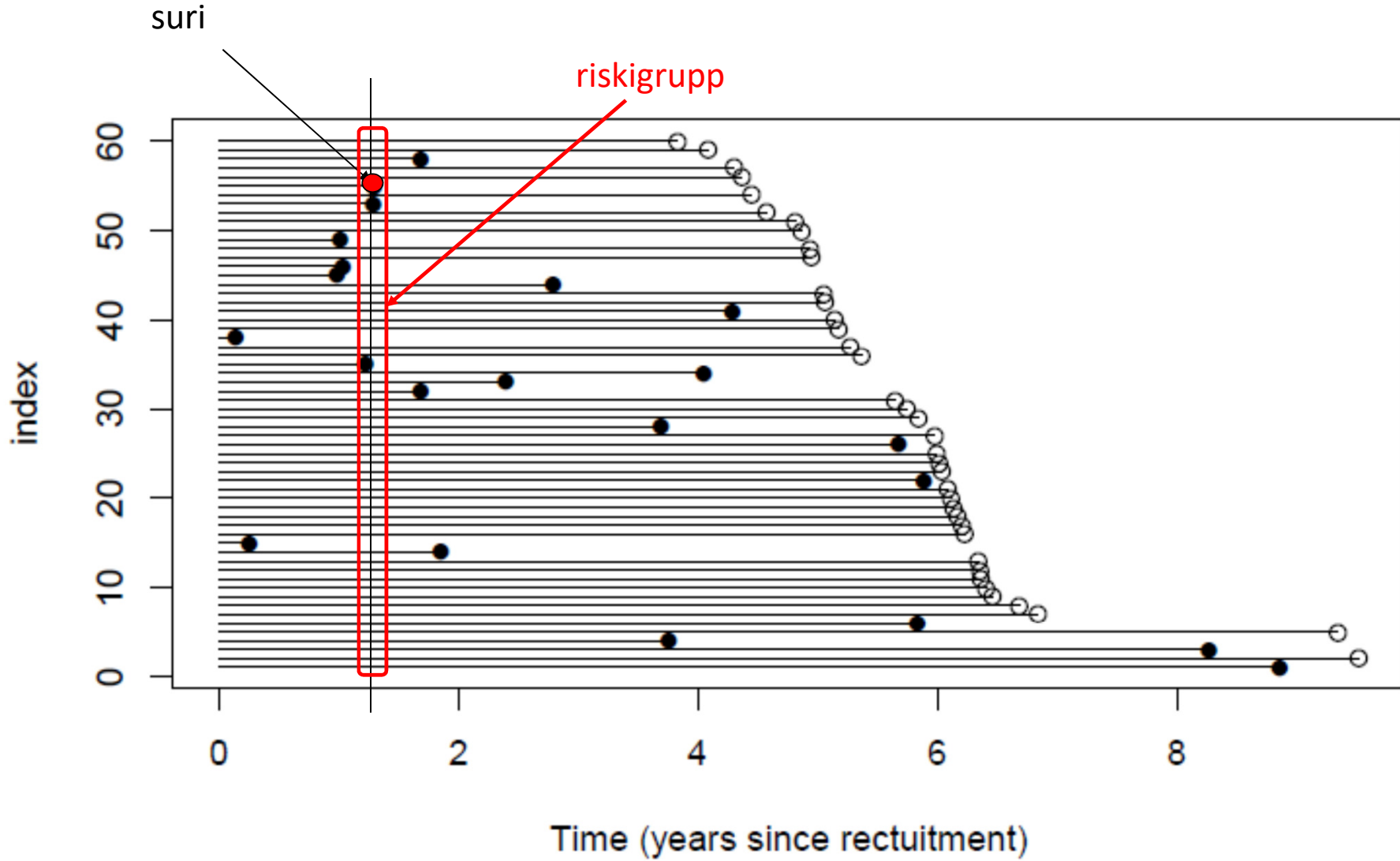
$$\psi(i) = e^{X_{1i}\beta_1 + \dots + X_{ki}\beta_k}$$

$$L = \prod_{j=1}^d \frac{\psi(j)}{\sum_{k \in R(\tau_j)} \psi(k)}$$

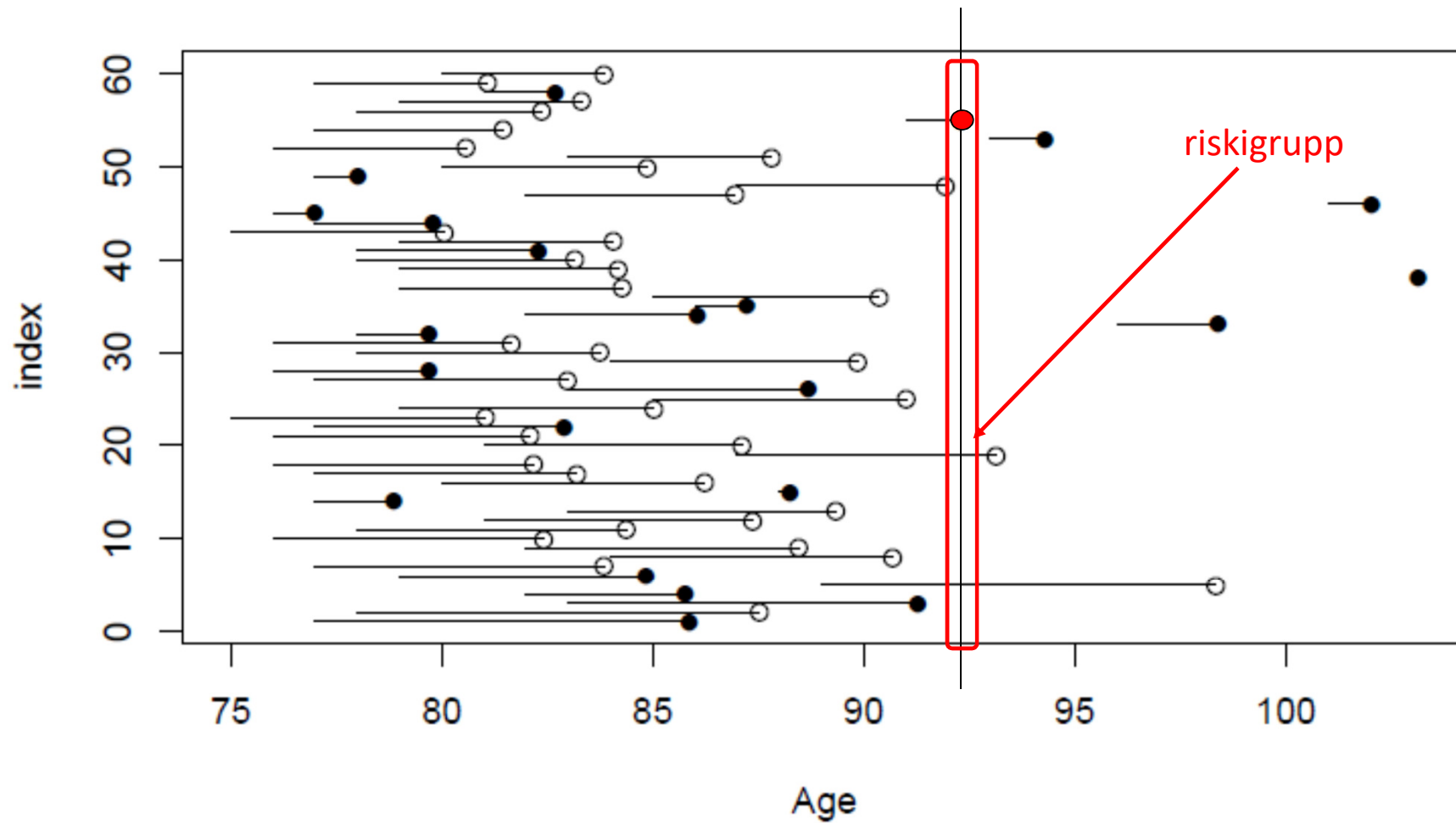
Kus j 's indiviid suri (haigestus) ajal τ_j ja $R(\tau_j)$ on indiviidide arv riskigrupis (*at risk*) ajal τ_j

Seega muutub erineva ajaskaala valiku korral riskigrupp $R(\tau_j)$ ja tekivad erinevad hinnangud

Tavaline valik ajaskaalaks: T – aeg liitumisest surmani

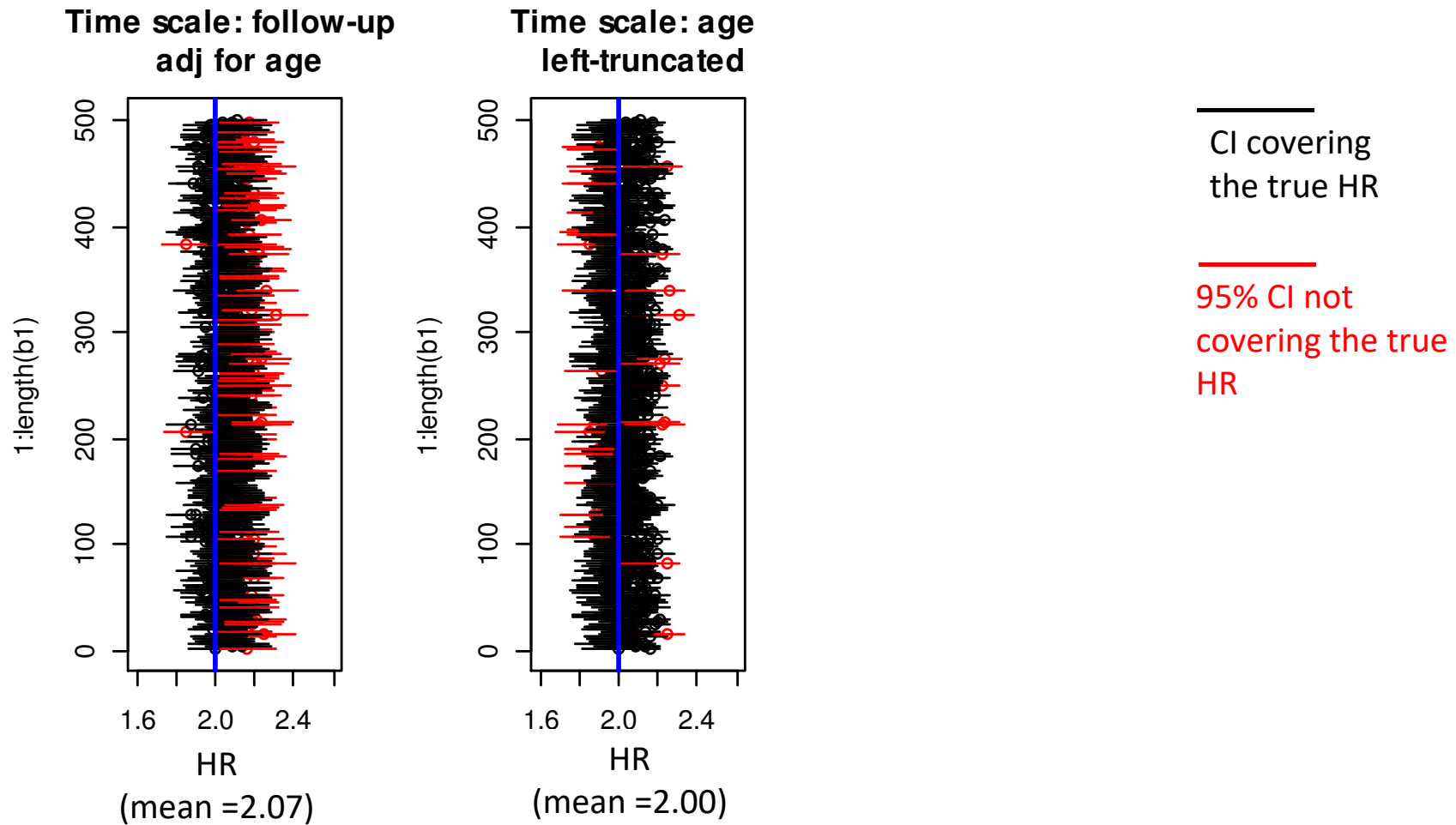


T - vanus

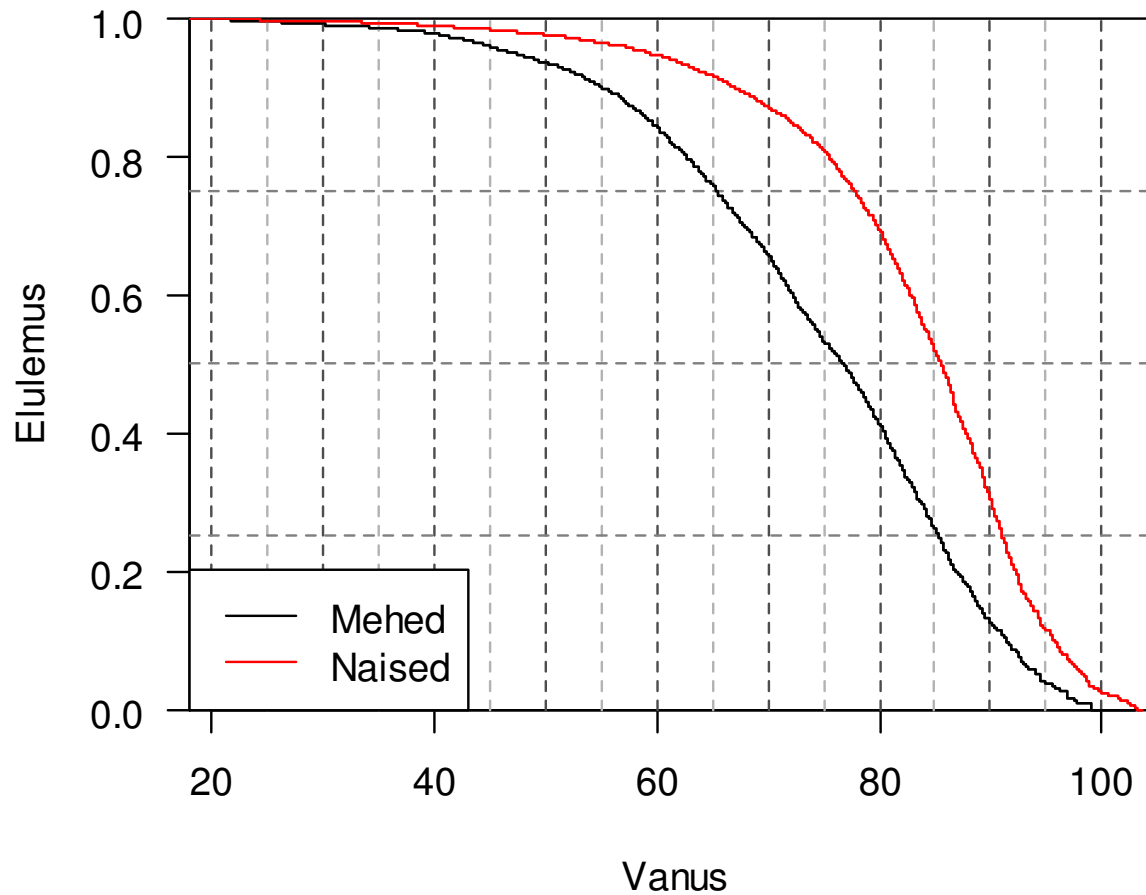


Simulatsioonide tulemusi.... (õige HR=2)

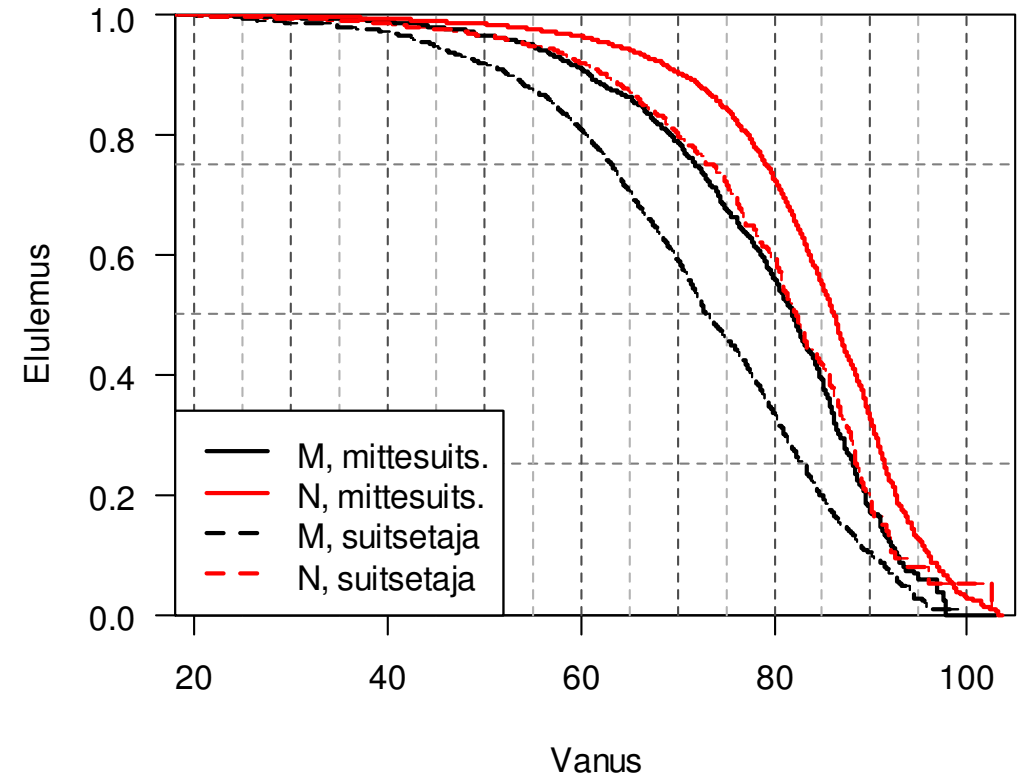
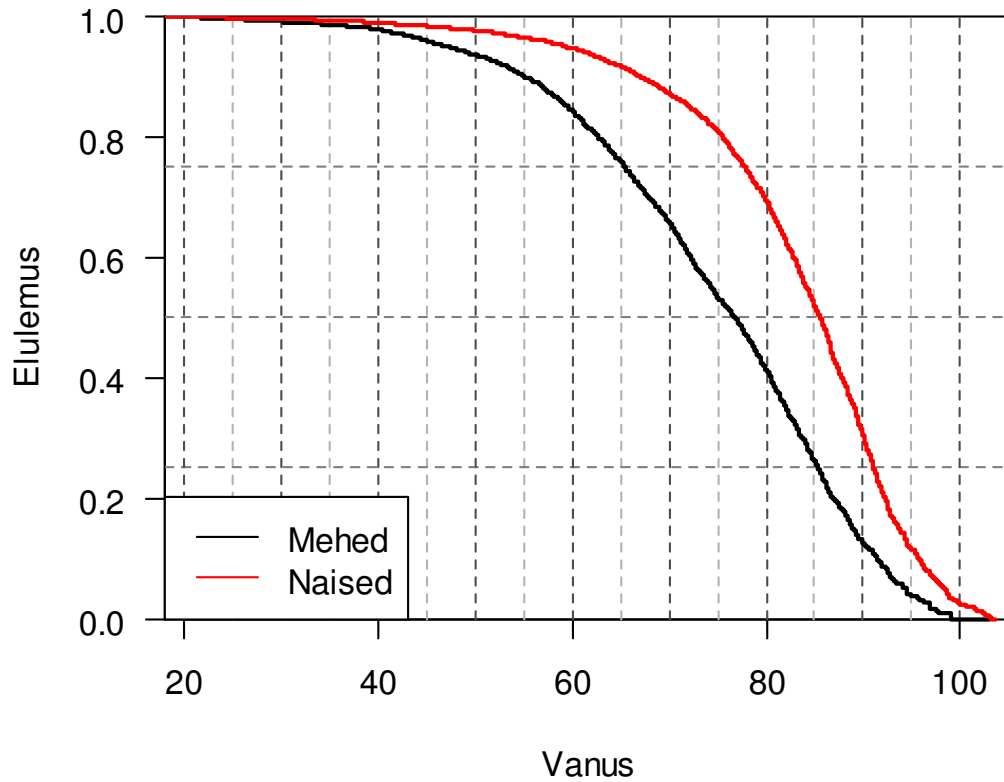
Hinnatud riskisuhted (95% CI), kasutades erinevaid ajaskaalasid



Eluea andmed TÜ Eesti Geenivaramus: elulemus (survival function) $P(T>t)$



Eluea andmed TÜ Eesti Geenivaramus: elulemus ja suitsetamine



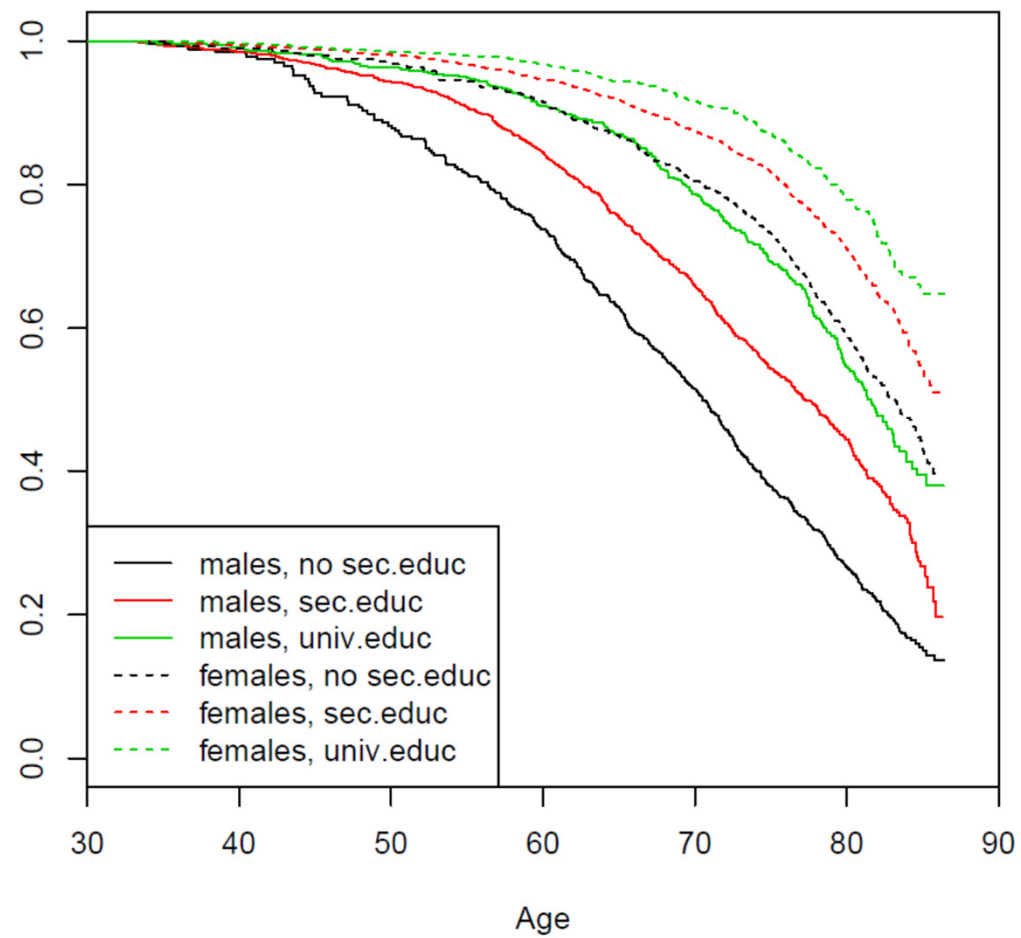
Olulisemad riskifaktorid

	RISKIMÄÄRADE SUHE (Hazard ratio)	95% CI
Sugu: mees	1.7	1.6-1.8
On suitsetanud	1.1	1.0-1.2
Suitsetab praegu	1.9	1.8-2.1
Keskhariduse puudumine	1.4	1.3-1.5
Kõrghariduse puudumine	1.3	1.2-1.5
Ei söö puuvilju	1.3	1.2-1.5
Rasvumine: KMI>35	1.3	1.2-1.4
Põeb vähki	1.7	1.6-1.9
On müok.infarkti läbi põdenud	2.0	1.8-2.1
I tüüpi diabeet	2.2	1.8-2.6
II tüüpi diabeet	1.6	1.5-1.7

Kui suures osas on need seosed põhjuslikud?

Hariduse mõjust veel...

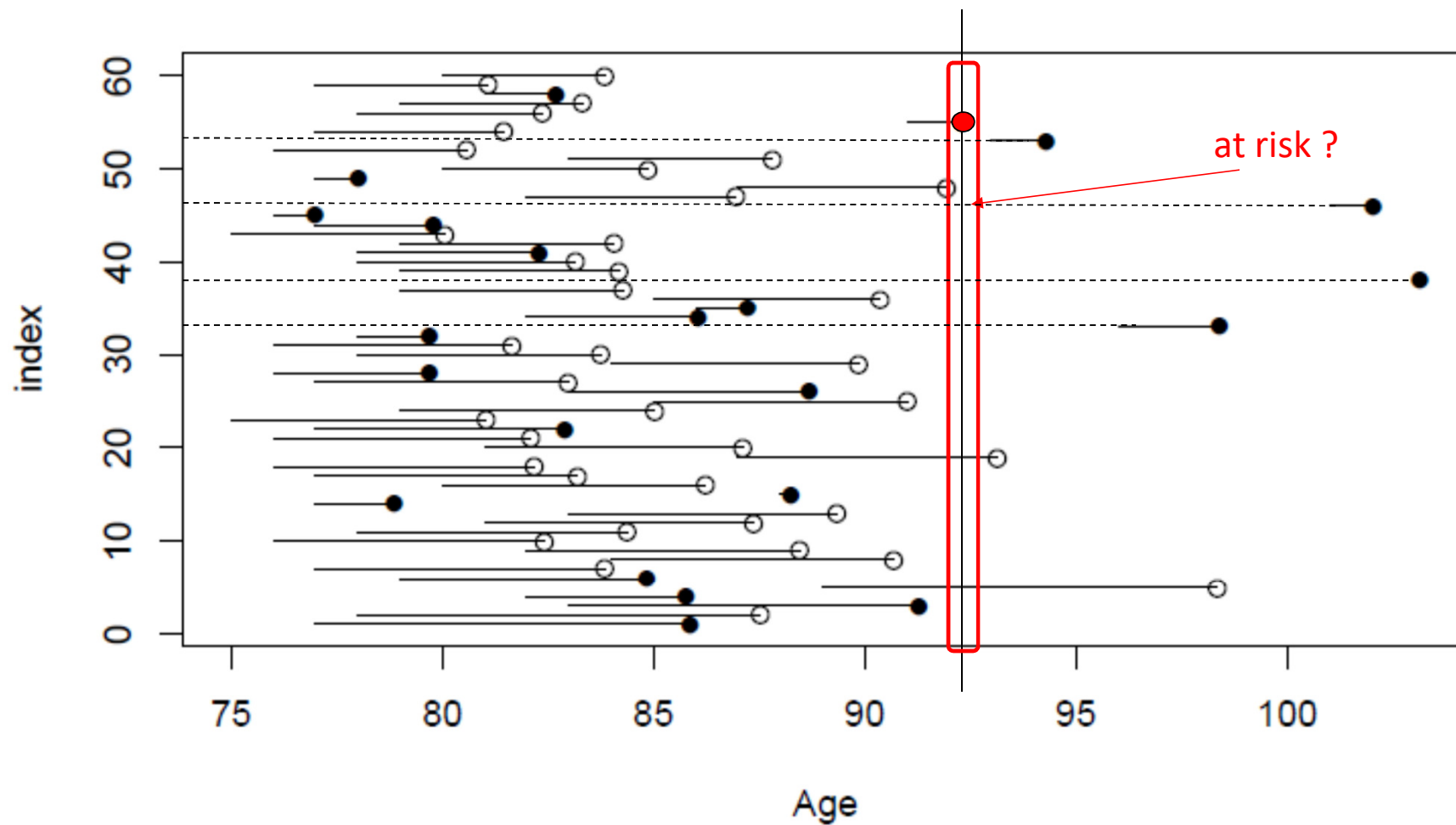
Survival in the Estonian Biobank cohort
(age >30, YOB>1930, n=36943, 2826 events)



Eluea geneetika?

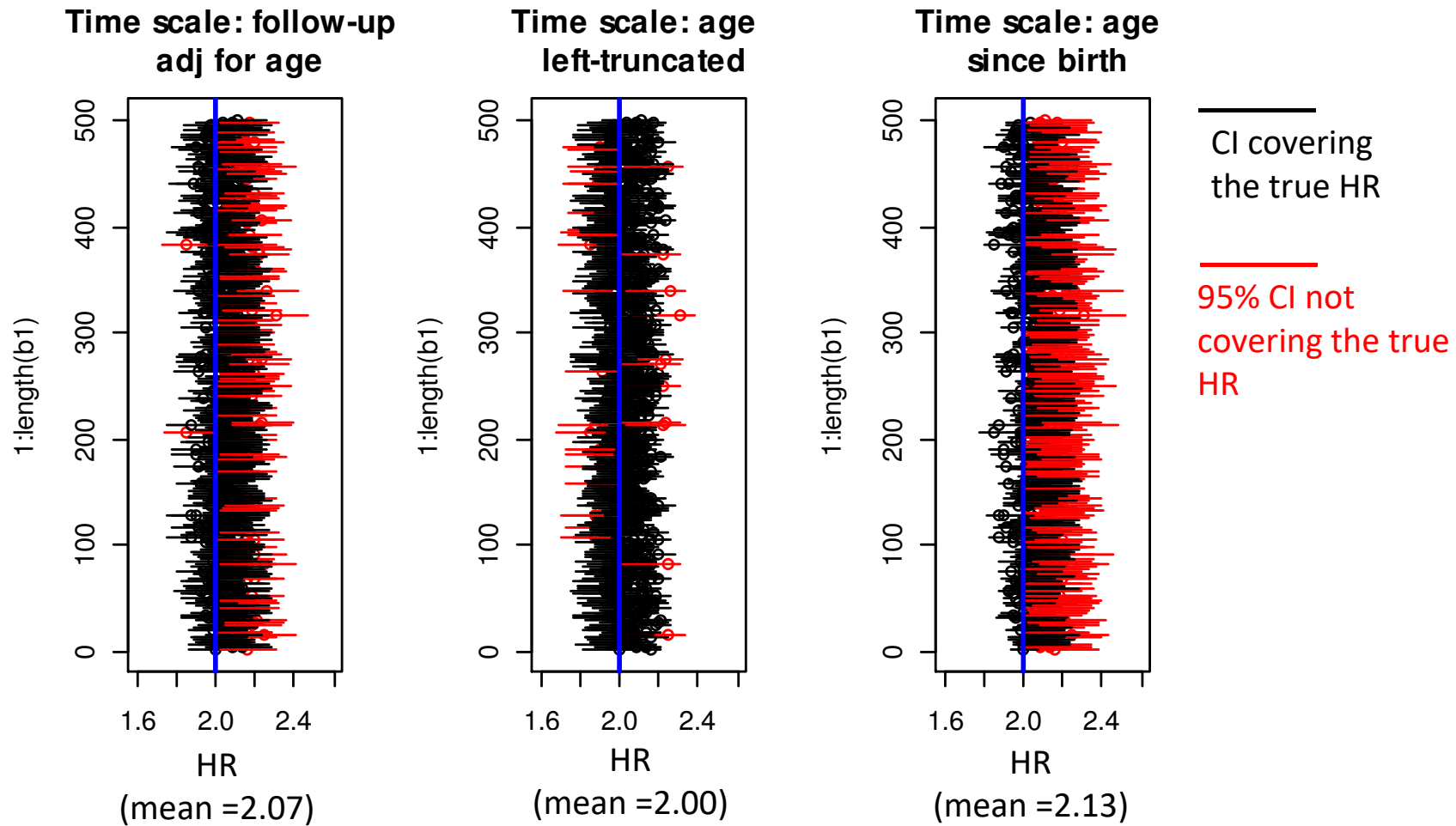
- Pikaeealisus eeldab madalat riski paljude haiguste suhtes – kas võib olla olemas mõni geen, mis seda soosib?
- Kaksikute-uuringute põhjal on pikaeealisuse päritavus 0.23-0.26
- Geeniuuringud andsid pikka aega vastukäivaid tulemusi – põhjuseks eelkõige väikesed valimimahud ja erinevused uuringukavandites
- Evolutsioon on tänaseks tekitanud olukorra, et võimalikud geenimõjud saavad olla vaid üsna nõrgad ja nende avastamiseks on vaja väga suuri valimimahtusid (mille määrab surnute arv valimis)!

Geenivariandid on fikseeritud juba sünnihetkel, kas me ei võiks alustada ajaskaalat sünnihetkest?



Simulatsioonide tulemusi.... (õige HR=2)

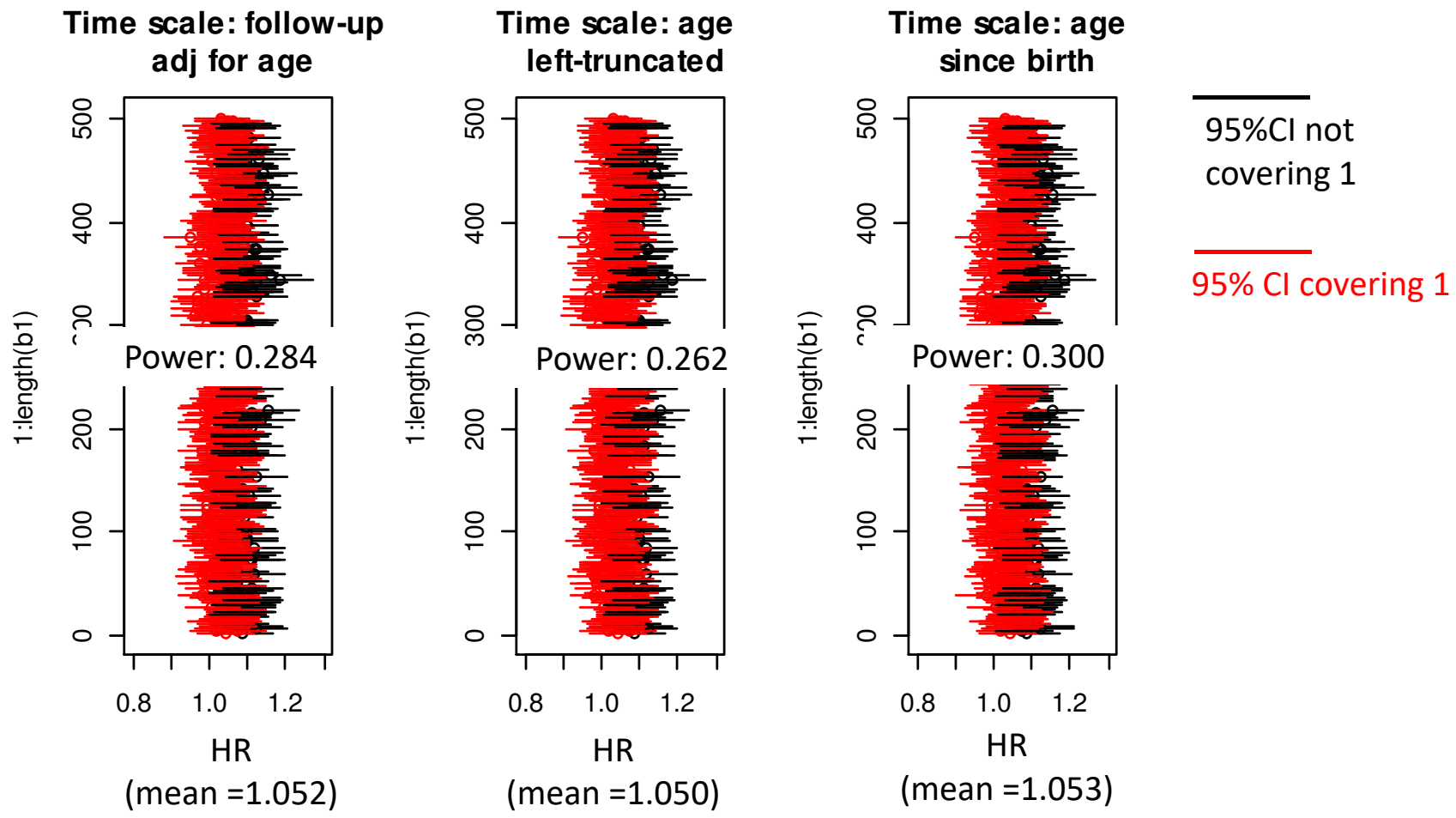
Hinnatud riskisuhted (95% CI), kasutades erinevaid ajaskaalasid



Kui me lisame riskigruppi need, kes antud vanuseks veel kohordis ei ole, tekitame nihke ülespoole!

Aga geenimõjud on väikesed – mis juhtub realistlikus olukorras, kus riskisuhe on väike (HR=1.05)?

Estimated Hazard Ratios (with 95% CI) using different time scales



„Vale“ ajaskaala korral tekib väike nihe, aga saame parima võimsuse!

Idee, kuidas veel võimsust suurendada: kasutame vanemate eluigasid indiviidi enda eluea asemel

Indiviidi genotüüp (ühenukleotiidsse polümorfismi ehk SNP osas) on kodeeritud kui kahe alleeli (0/1) summa:

$$X = a_1 + a_2$$

a_1 on saadud ühelt ja a_2 teiselt vanemalt

Vanemate genotüübid X_1 ja X_2 koosnevad alleelist, mis on lapsele edasi antud ja teisest alleelist:

$$X_1 = a_1 + b_1 \quad X_2 = a_2 + b_2$$

Kasutame vanemate eluigasid (T_1 ja T_2), aga järglase genotüüpi X (X_1 ja X_2 asemel)

Seega $\text{cor}(X, X_1) = \text{cor}(X, X_2) = 0.5$ ja vastavad mudeli parameetrid on 2 korda väiksemad.

(HR = ruutjuur „õigest“ HR-st)

Geneetilistest mõjudest: 1. ülegenoomne analüüs UKBB andmete põhjal, kasutades vanemate eluea andmeid

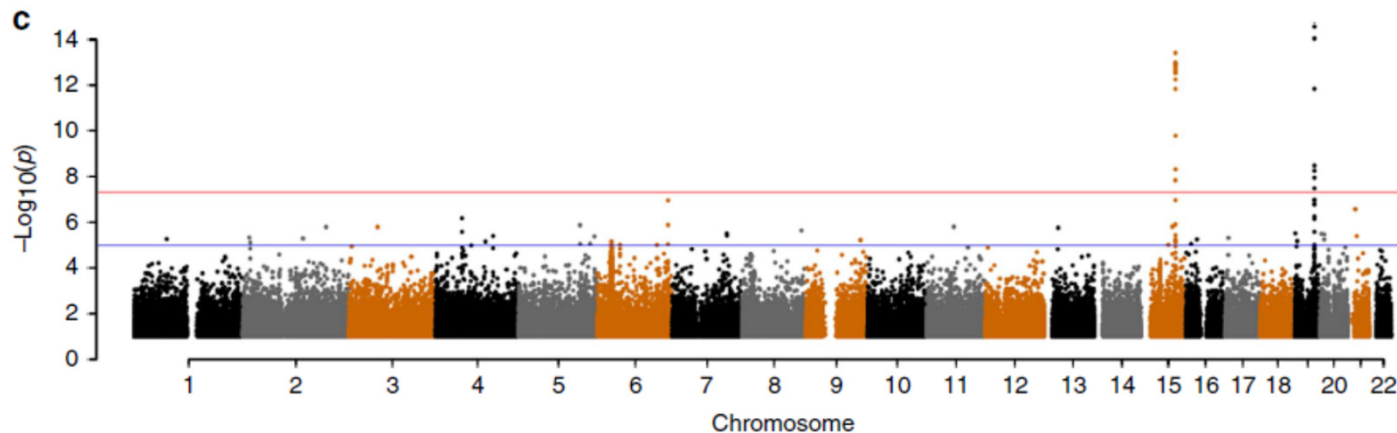
Received 28 Sep 2015 | Accepted 29 Feb 2016 | Published 31 Mar 2016

DOI: 10.1038/ncomms11174

OPEN

Variants near *CHRNA3/5* and *APOE* have age- and sex-related effects on human lifespan

Peter K. Joshi¹, Krista Fischer², Katharina E. Schraut^{1,3}, Harry Campbell¹, Tõnu Esko^{2,4,5,6} & James F. Wilson^{1,7}



>270000 **parental** lifespans from the UK Biobank, linear regression for Martingale residuals

Geneetilistest mõjudest: 2. ülegenoomne analüüs täienenud UKBB + veel mõne Euroopa kohordi andmete põhjal, kasutades vanemate eluea andmeid



ARTICLE

DOI: [10.1038/s41467-017-00934-5](https://doi.org/10.1038/s41467-017-00934-5)

OPEN

Genome-wide meta-analysis associates *HLA-DQA1/DRB1* and *LPA* and lifestyle factors with human longevity

Peter K. Joshi et al. 

[Nat Commun.](https://doi.org/10.1038/s41467-017-00934-5) 2017 Oct 13;8(1):910. doi: 10.1038/s41467-017-00934-5.

Eesti kaasautorid: Krista Fischer, Tõnu Esko, Lili Milani, Reedik Mägi

Nüüd õnnestus valimit suurendada >600000-ni

- Idea: use parental lifespans instead of observed lifespans in the cohort(s)
- Data: UK Biobank (European and African ancestry populations) + 24 other cohorts (mainly European)

Table 1 Summary of the LifeGen parental lifespans

Ancestry	Parent	Count		
		Alive	Dead	Total
African	Father	2435	6924	9359
African	Mother	4185	5889	10,074
European	Father	113,611	178,017	291,628
European	Mother	150,854	144,144	294,998
	ALL	271,085	334,974	606,059

Nüüd kokku 4 geeni....

Table 2 Four regions associated with lifespan at genome-wide significance and replicatio

rsid	Gene	a1	Freq a1	N(000) parent	HR a1	SE	P-value	Years
rs34831921	* <i>HLA-DQA1 /DRB1</i>	A	0.09	481	0.942	0.011	4.18 E-08	0.6
rs55730499	* <i>LPA</i>	T	0.083	563	1.074	0.011	8.67 E-11	-0.7
rs8042849	<i>CHRNA3/5</i>	C	0.356	567	1.046	0.006	3.75 E-14	-0.4
rs429358	<i>APOE</i>	C	0.142	556	1.091	0.008	1.44 E-27	-0.9

*new

HLA: Known associations with rheumatoid arthritis and Crohn's disease

LPA: Associations with blood lipids and coronary traits

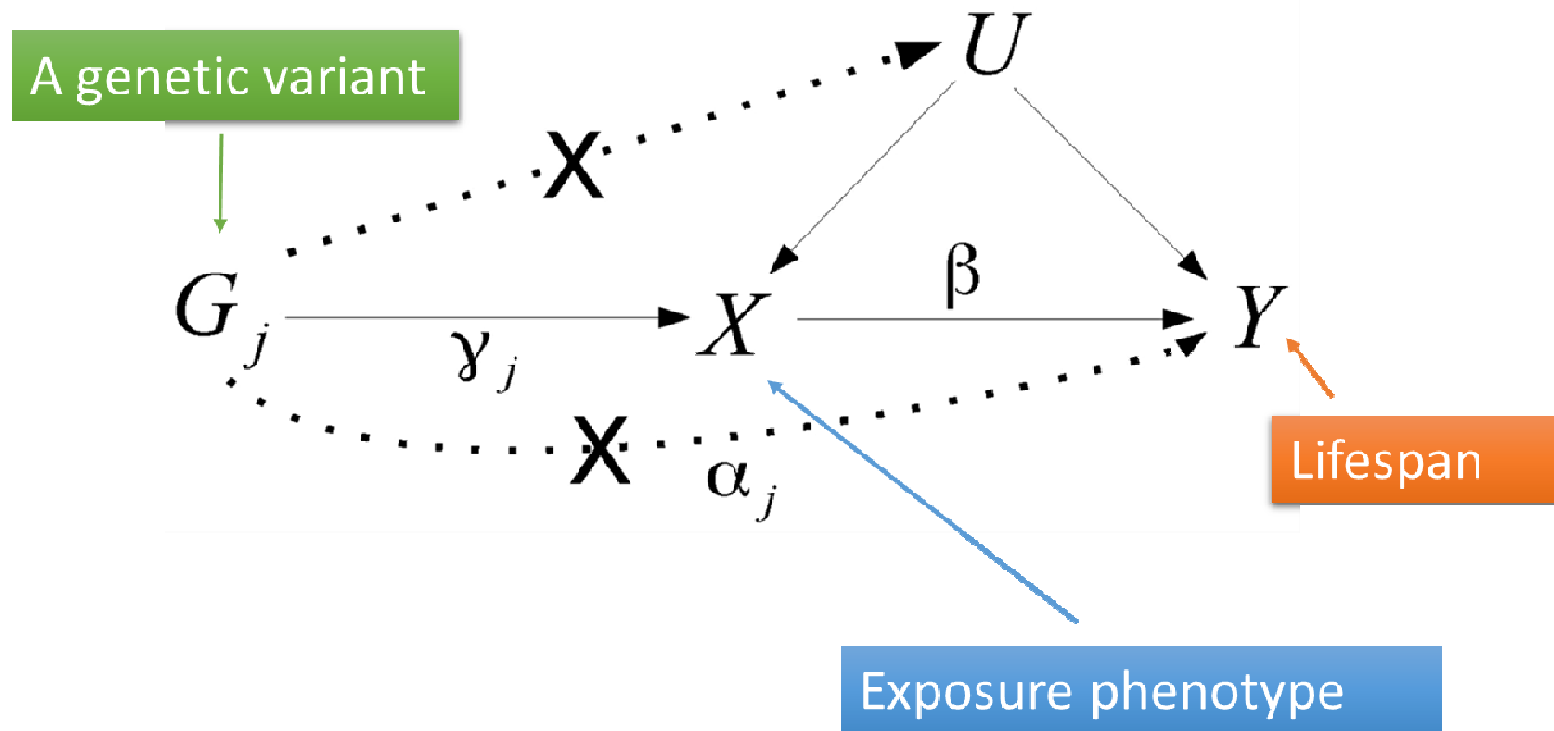
CHRNA3/5: associations with smoking behavior

APOE: associations with Alzheimer's disease, blood lipids, cardiac and cognitive ageing traits

Mendelian randomization: causal relationships with lifespan

Põhjuslikud seosed elueaga

- Using MRbase and available summary statistics for several traits
- Two-sample Mendelian Randomization to investigate causal influences on lifespan – tested more than 90 phenotypes, seven appeared to be significant



**Interquartile effect
in years**

<i>Risk factor</i>		
Body mass index SD	65	3.8
(kg/m ²)		
Years of schooling	64	-4.7
SD (years)		
Cigarettes smoked	rs12914385	5.3
per day		
HDL cholesterol SD	39	-1.4
(mg/dL)		
LDL cholesterol SD	17	1.4
(mg/dL)		
Fasting insulin log	6	4.1
pmol/L		
SBP mmHg	rs381815	5.2
CRP log mg/L	39	-0.66

Reduced life: 7 months per BMI unit

Increases lifespan: 11 months per year of education

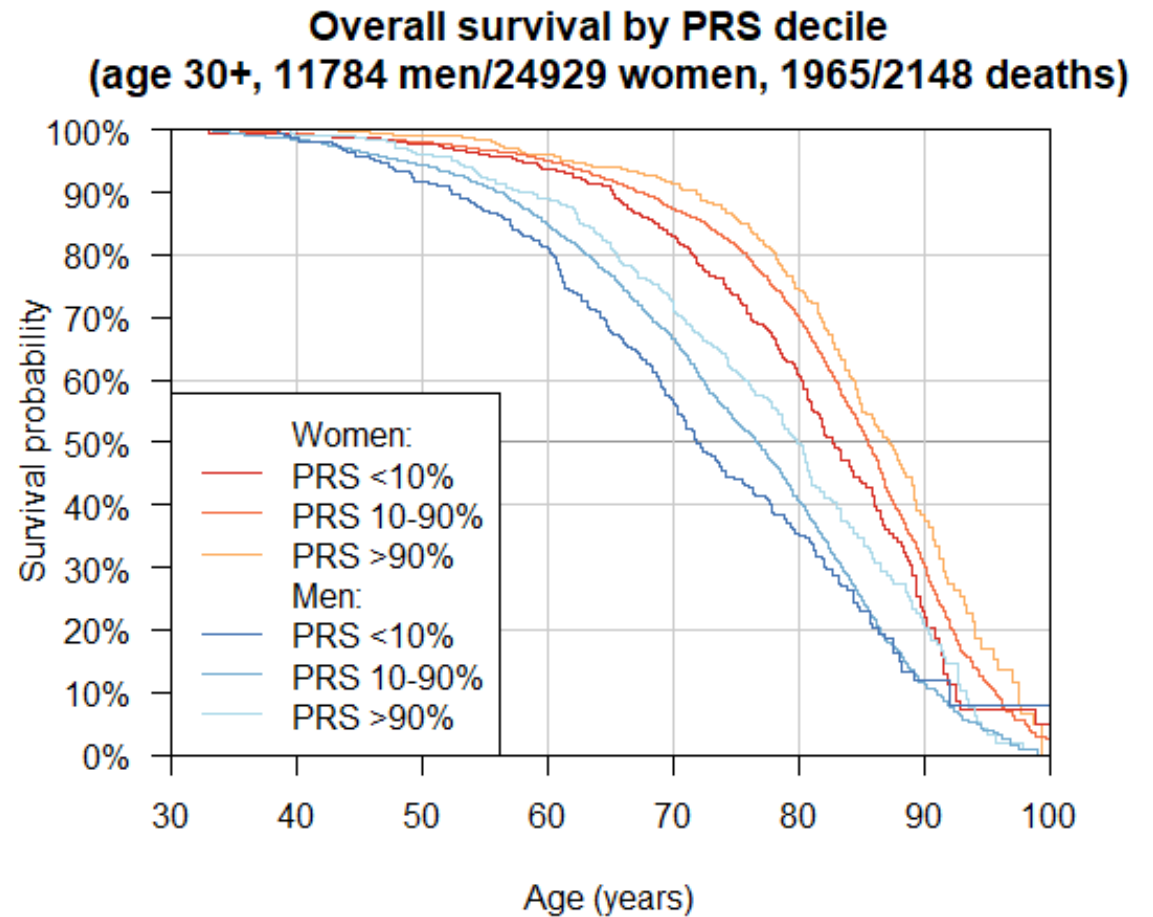
Lifelong smoking of 20 cigarettes a day reduces lifespan by 6.8 years

Increased CRP increases lifespan!

Viimased tulemused (Timmers et al, bioRxiv)

<https://www.biorxiv.org/content/early/2018/07/06/363036>

- Elukestuse geneetilise skoori



Kokkuvõttes: mis siis ikka mõjutab eluiga?

- Kõige tugevamalt mõjutavad eluiga eluviisid ja elukeskkond: suitsetamine, haridus, jms
- Samal ajal on eluviiside valiku juures olulisel kohal geenide mõjud
- *...aga geneetiline eelsoodumus suitsetamiseks ei sunni veel suitsetama?*

Aitäh!

Kristi Läll (TÜ Matemaatika ja statistika instituut, TÜ Genoomika instituut)

Tõnu Esko (TÜ Genoomika instituut)

Peter Joshi, Jim Wilson

(MRC Institute of Genetics and Molecular Medicine, Edinburgh, UK)

Zoltan Kutalik

(University of Lausanne, Switzerland)



TARTU ÜLIKOOL

Kodukootud fülogeneetiline meta-analüüsi mudel

Ants Kaasik

TÜ Ökoloogia ja Maateaduste Instituut

23.03.2018

Narva



TARTU ÜLIKOOL Lõpptulemus

Global Change Biology / Volume 24, Issue 4

RESEARCH REVIEW

A review of urban impacts on avian life-history evolution: Does city living lead to slower pace of life?

Tuul Sepp✉, Kevin J. McGraw, Ants Kaasik, Mathieu Giraudeau

First published: 23 November 2017

<https://doi.org/10.1111/gcb.13969>



Cinclus cinclus
Vesipapp



Andmed

Võrdleme lindude elu maal ja linnas

kurna suurus (25 andmepunkti)

elukestus (21 andmepunkti)



*Corvus
brachyrhynchos*
Ameerika vares



Täpsushinnangud

võimalusel arvatatud

kohati ennustatud valimimahu põhjal



Falco naumanni
Tuuletallaja



Andmed (3)

Mitu andmepunkti samast uuringust

Uuring on juhuslik efekt

Mõned linnuliigid on lähemalt seotud kui teised (fülogenees)

Liik on korreleeritud juhuslik efekt



Falco tinnunculus
Harilik tuuletallaja



Uuringute erinev kestus

Juhuslik efekt hõlmab endas nn aasta mõju

Kui meil on mitmete aastate keskmine siis dispersioon ei saa olla sama



Fulica Atra
Lauk

Kokkuvõttes

efekti suurus =

üldkeskmine +

kovariaatide mõju +

liigi mõju +

uuringu mõju

Vaadeldud $N(\text{efekti suurus}, SE)$

Aga uuringu mõju on kujul

$$N\left(0, \frac{\sigma}{\sqrt{t}}\right)$$



Parus caeruleus
Sinitihane



Analüüs?



Traditsiooniline segamudelite (või meta-analüüsi) tarkvara on sellises olukorras võimetu

AGA



Murrame siis traditsioone!



Parus major
Rasvatihane



Stan

*StanCon 2018 California is over, see you in StanCon
2018 Helsinki, Aug 29-31*

Stan^o is a state-of-the-art platform for statistical modeling and high-performance statistical computation. Thousands of users rely on Stan for statistical modeling, data analysis, and prediction in the social, biological, and physical sciences, engineering, and business.

Users specify log density functions in Stan's probabilistic programming language and get:

- full Bayesian statistical inference with MCMC sampling (NUTS, HMC)
- approximate Bayesian inference with variational inference (ADVI)
- penalized maximum likelihood estimation with optimization (L-BFGS)

Stan's math library provides differentiable probability functions & linear algebra (C++ autodiff). Additional R packages provide expression-based linear modeling, posterior visualization, and leave-one-out cross-validation.



Passer domesticus
Koduvarblane


```

46
47 transformed parameters{
48     vector[N] keskm;
49     matrix[K1,K1] T;
50     keskm=beta_1 + (X2*studyef)./sqrt(studyyears)+X1*speciefef;
51     T=pow(sigma_species,2)*treecorr;
52
53 }
54 model {
55     beta_1~cauchy(0,mu_sd_prior);
56     sigma_study~cauchy(0,cauchy_sd_prior);
57     sigma_species~cauchy(0,cauchy_sd_prior);
58     studyef~normal(0,sigma_study);
59     speciefef~multi_normal(Zero, T);
60     clutch~normal(keskm,SE);
61
62 }
63

```



Pica pica
Harakas

Tulemused

TABLE 1 Meta-analysis on the urbanization effects on survival and clutch size in birds by fitting a Gaussian mixed model in a Bayesian framework. Full distribution of the annual survival probability and clutch size models

	Mean	SE mean	SD	2.5%	5%	25%	50%	75%	95%	97.5%	N eff	Rhat
Survival												
Beta_1	0.142	0.005	0.138	-0.130	-0.069	0.060	0.145	0.226	0.357	0.408	838.086	1.001
Beta (sex)	0.007	0.000	0.033	-0.057	-0.047	-0.016	0.006	0.029	0.062	0.072	4580.222	1.000
Beta (method)	0.138	0.003	0.175	-0.210	-0.152	0.030	0.137	0.250	0.422	0.490	2627.082	1.000
Sigma (study)	0.109	0.005	0.103	0.007	0.011	0.038	0.078	0.145	0.314	0.391	447.781	1.007
Sigma (species)	0.203	0.002	0.064	0.109	0.118	0.158	0.193	0.235	0.320	0.357	1518.503	1.000
Clutch size												
Beta_1	0.358	0.008	0.336	-0.343	-0.211	0.155	0.367	0.575	0.880	0.998	1836.345	1.000
Sigma (study)	1.200	0.004	0.242	0.811	0.855	1.029	1.172	1.340	1.639	1.755	3853.107	1.002
Sigma (species)	0.600	0.007	0.259	0.242	0.276	0.411	0.548	0.737	1.098	1.233	1379.352	1.001



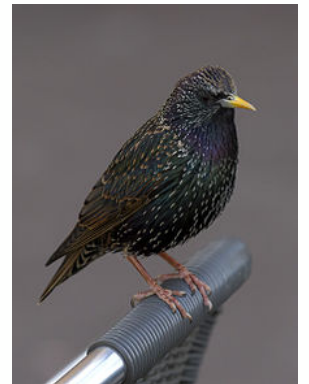
Vanellus miles
Lokutkiivitaja



TARTU ÜLIKOOL

Aitäh kuulajatele!

Linnupildid: Wikipedia



Sturnus vulgaris
Kuldnokk

Elurikkuse andmed ja andmehaldus

Urmas Kõljalg
Tartu Ülikool



Elurikkuse andmete eripära
Andmehaldus
Masinloetavad FAIR andmed
Rakendused

Elurikkuse andmete eripära

Liigi (taksoni) nimi elurikkuse andmete tähtsaim kommunikatsiooni vahend.

Elurikkuse andmed pea alati seotud taksoni nime ja klassifikatsiooniga

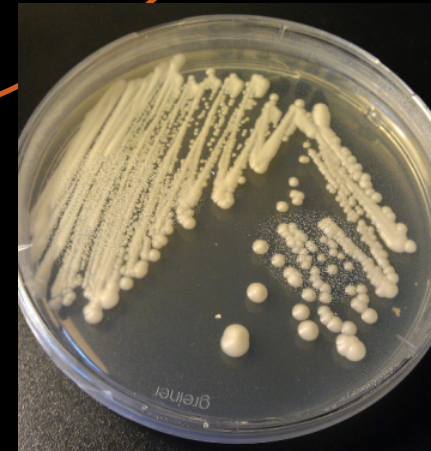
Taksoni eristamine ja tema nimetamine.

Rana temporaria Linnaeus, 1758



Taksoni eristamine ja tema nimetamine.

Saccharomyces cerevisiae

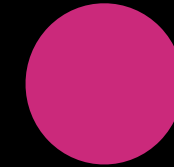


Eluslooduse klassifikatsioonid

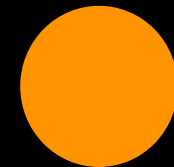
UNITE

GenBank

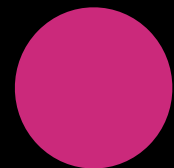
● Kingdom name



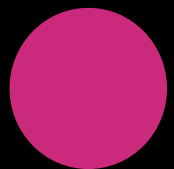
● Phylum name



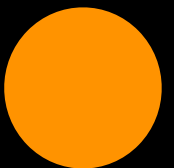
● Class name



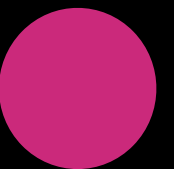
● Order name



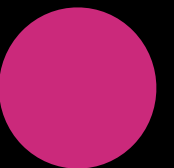
● Family name



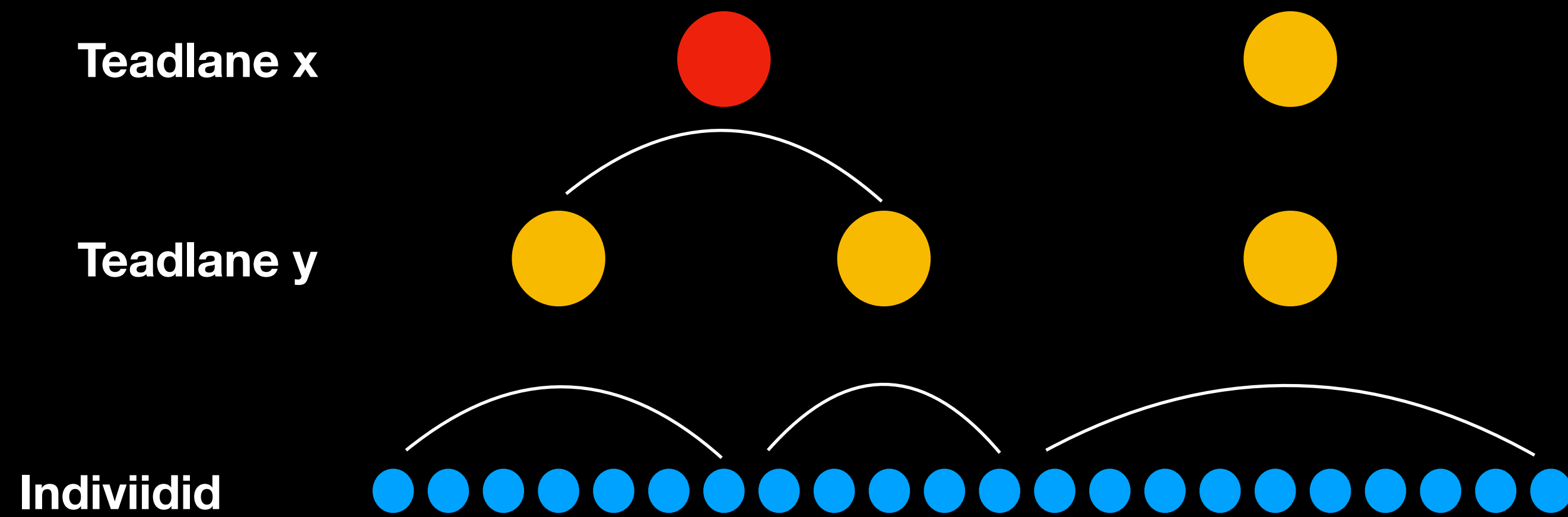
● Genus name



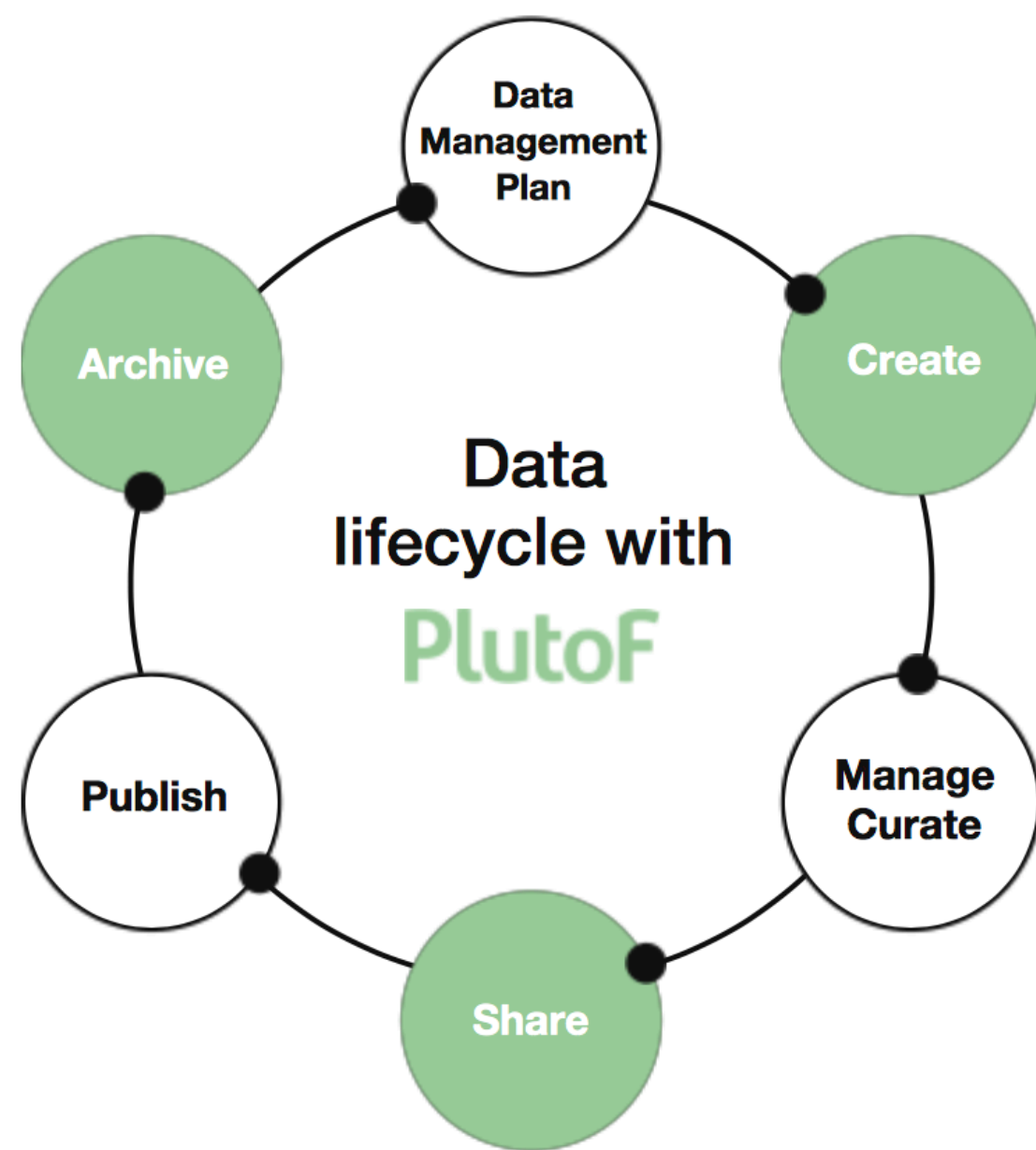
● Species name



Liikide eristamine



Andmehaldus



Data Management and Publishing Platform

In PlutoF you can manage your full data lifecycle. From Data Management Plan to the publishing and archiving your datasets in machine readable format.

Workbench Modules



Projects

Manage all your projects online and connect them to different data and datasets. Every project may include unlimited number of localities and associated data. It may be public, private or shared with selected workgroups.



Collections

Biological collections may create and manage their digital archives. There are specific modules for the printing of labels, managing of loans, compiling of reports, publishing of datasets in GBIF or other places. Types of the collection datasets include preserved or living specimens and environmental samples.



Monitoring and Conservation

PlutoF may also help in creating and managing your projects of monitoring and nature conservation. Single project may combine different records of taxon occurrences, human and DNA based observations, samples, specimens, references, etc. Datasets



File Repository

Most data may easily be linked to the files which are uploaded and managed through the file repository module. For example, images or videos may be linked to the sampling area, observation, specimen, reference, institution, person, project, etc. Sound, raw DNA sequences, digitized notebooks are also good examples of the files managed through this module.



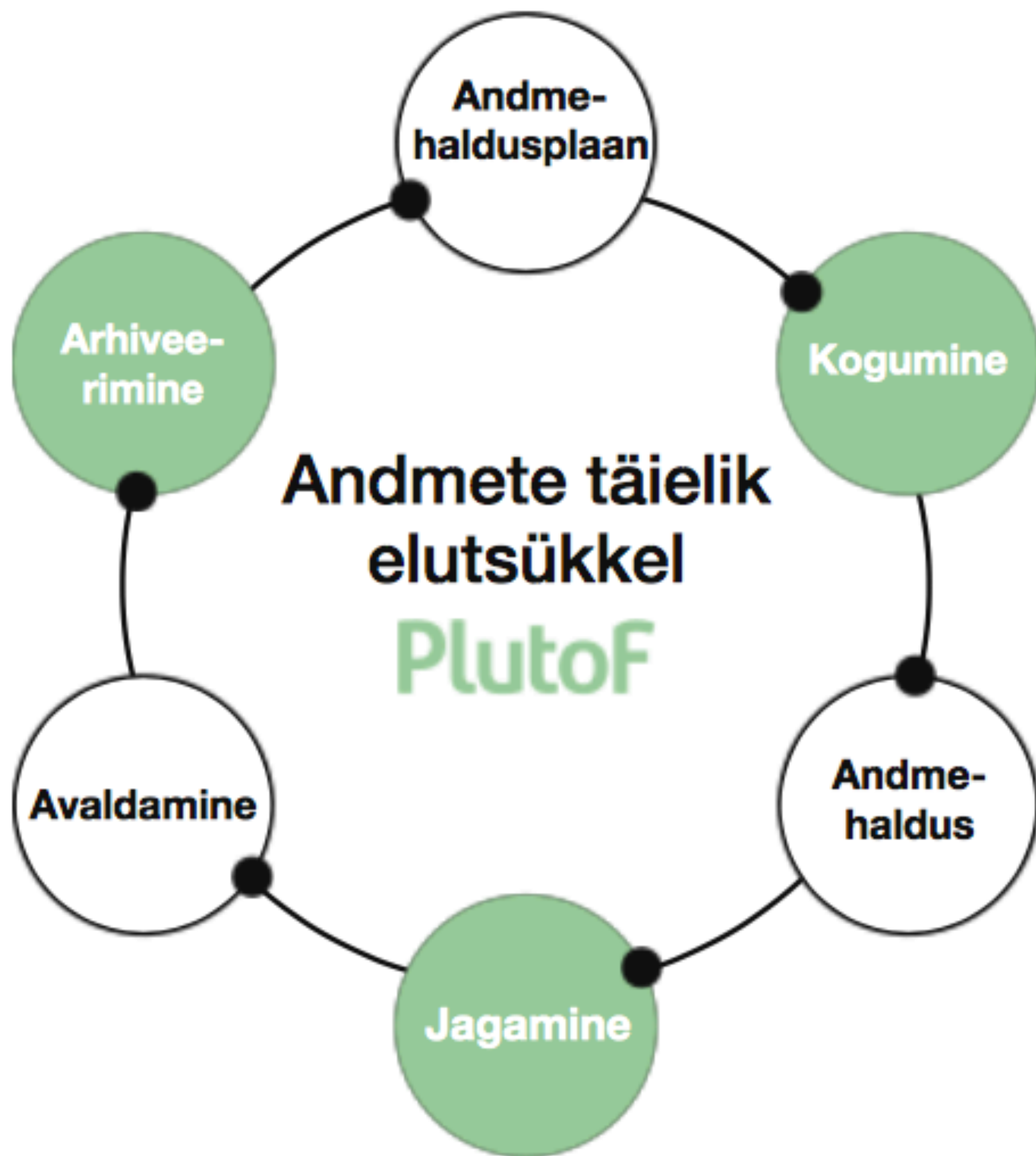
Analyses

Specific analyses can be run online and the number of analyses accessible is growing. Examples of current analyses include sequence analyses with different programs (e.g. ITSx, ATOSH, massBLASter).



Publishing

Here you can ask for a new Digital Object Identifier (DOI) for your dataset. Datasets may be published via GBIF also. Data of references and keywords is managed



Masinloetavad FAIR andmed

Open Data (avaandmed)
&

FAIR DATA (õiglased andmed)

Findable, Accessible, Interoperable and Re-usable

Masinloetavad FAIR andmed

Andmete publitseerimine PlutoFis

DOI (Digital Object Identifier)

Publitseeritud enam kui pool miljonit DOI põhist andmekogumit

Rakendus



eElurikkus koondab ühte portaali Eesti looduse andmed.
Juba on talletatud 28 248 liiki

LIIGID

Loomad Seened Taimed Protistid Bakterid

ASUKOHT

Otsi hulknurga abil Avasta ümbruskonnas Eeldefineeritud alad

LIIGI ESINEMISED

Täpsem otsing Mitme taksoni nime otsing Kataloogi numברי otsing

Monitooring

Päevaliblikate kooslused
Ööliblikate kooslused

DNA põhiseid vaatlused

UNITE andmeressurs

Loodusvaatlused

Viimased vaatlused
Viimased linnuvaatlused

Haudelindude levikuaatlas

Eesti taimede levikuaatlas

Avasta ümbruskonnast

Avasta või täienda Eesti liigirikkust huvi pakkuvast piirkonnast. Otsi liigi esinemisi aadressi, kohanime, posti sihtnumbri või GPS koordinaadi (laiuskraad, pikkuskraad) järgi. Kõik kirjed on majutatud ning hallatavad [PlutoF](#) infosüsteemist. Ootame ka sinu tähelepanekud Eesti loodusest.

[Regioonid](#) [Geograafiline otsing](#)

Narva kolledz

Tulemused asukohas: **Raekoja plats 2, 20307 Narva, Estonia**

Liigirühm	#	Liik : Tavanimetus	Kirjeid
➤ Kõik liigid	672	1. <i>Acanthocinus aedilis</i> : Harilik käätsusikk	1
➤ Loomad	0	2. <i>Acanthopsyche atra</i> : ogamärslane	2
		3. <i>Acleris hastiana</i>	1
		4. <i>Acleris logiana</i>	1
		5. <i>Acronicta (Triaena) tridens</i> : kolmkida-noolöölane	1
		6. <i>Acronicta (Viminia) menyanthidis</i> : soo-noolöölane	1
		7. <i>Aglais urticae</i> : koerliblikas	2
		8. <i>Agrotis clavis</i> : korpjas mullaöölane	2
		9. <i>Agrotis segetum</i> : oraseöölane	1
		10. <i>Agrotis vestigialis</i> : männitõusmeöölane	4
		11. <i>Amara (Amara) communis</i>	1
		12. <i>Amara (Amara) nitida</i>	1
		13. <i>Amphipoea fucosa</i> : tavaline segaöölane	2
➤ Linnud	70	14. <i>Amphipoea lucens</i> : soo-segaöölane	2
		15. <i>Amphipyra perflua</i> : haava-võsaöölane	1
		16. <i>Amphipyra tragopoginis</i> : väike-võsaöölane	1
		17. <i>Anarta trifolii</i> : ristikuöölane	3
		18. <i>Anisodactylus (Anisodactylus) binotatus</i> : punalaik-ehmesjooksik	1
		19. <i>Anoplodera virens</i> : Rohekas-õiesikk	2
		20. <i>Anthocharis cardamines</i> : koiduliblikas	1

Radius (km)





Unified system for the DNA based fungal species linked to the classification Ver. 7.2



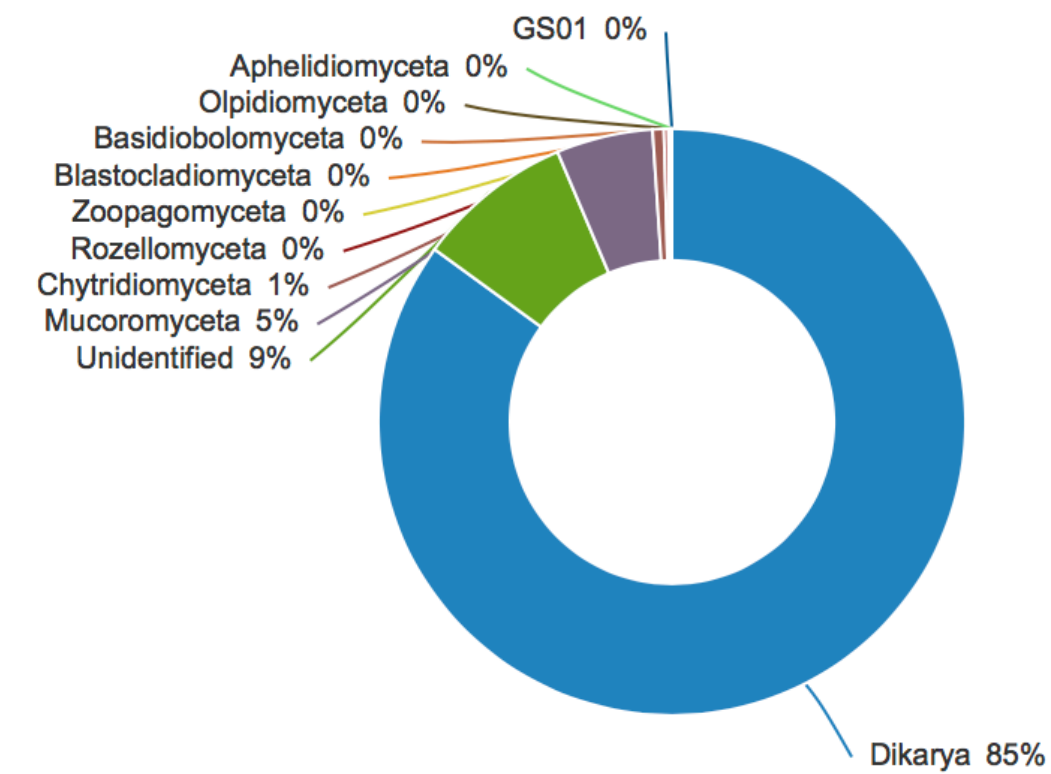
Current version: 7.2; Last updated: 2017-12-01 (read more)

Number of ITS sequences (UNITE+INSD): 817 130; Number of UNITE fungal Species Hypotheses with DOIs at 1.5% threshold: 73 929 (more statistics)

Threshold: 1.0 % Include: All SH-s Start typing taxon name here ... Go Reset ⓘ

- ▶ [Dikarya \(75,999\)](#)
- ▶ [Unidentified \(7,696\)](#)
- ▶ [Mucoromyceta \(4,750\)](#)
- ▶ [Chytridiomyceta \(553\)](#)
- ▶ [Rozellomyceta \(211\)](#)
- ▶ [Zoopagomyceta \(85\)](#)
- ▶ [Blastocladiomyceta \(43\)](#)
- ▶ [Basidiobolomyceta \(26\)](#)
- ▶ [Olpidiomyceta \(16\)](#)
- ▶ [Aphelidiomyceta \(5\)](#)
- ▶ [GS01 \(informal name, see Tedersoo et al., 2017\) \(2\)](#)

SH graph: Fungi



Tomentella badia | SH177891.07FU | DOI: [10.15156/BIO/SH177891.07FU](https://doi.org/10.15156/BIO/SH177891.07FU)

Distance to the closest SH: 1.5

No. of sequences in SH: 5

Placement in the fungal classification

Fungi; Basidiomycota; Agaricomycotina; Agaricomycetes; Agaricomycetes sbc Incertae sedis; Thelephorales; Thelephoraceae; Tomentella

Reference sequence: UDB000238|AF272937|UK91

Chosen by: Urmas Kõljalg

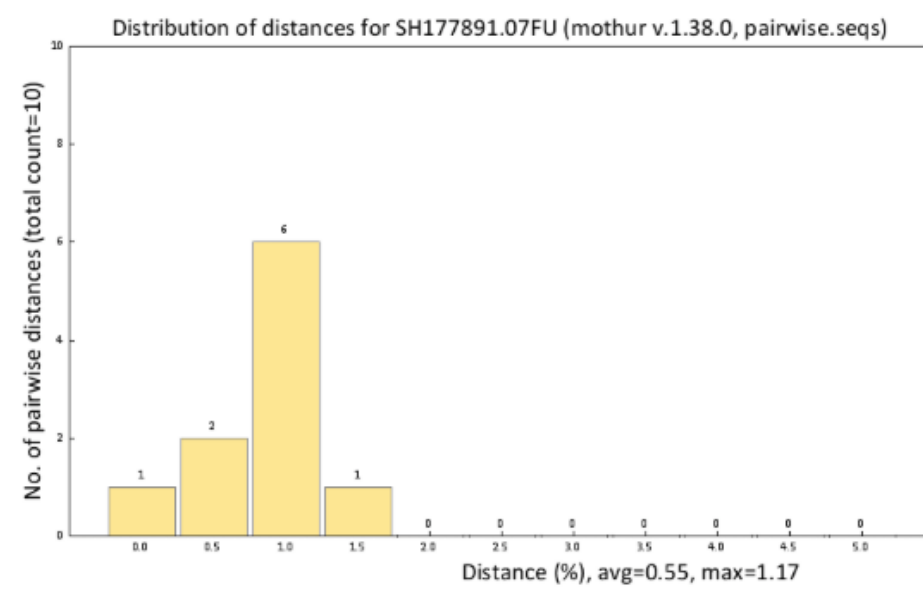
Date: 2014-11-09 22:05

Statistics

Taxonomy

Ecology

Distribution of distances



Older version(s) of this SH is/are available

SH code (Count*/Total count**)

[SH202603.06FU](#) (3/5);

*Number of sequences carried over from version 6.0

**Total number of sequences composing this SH in version 7.0

Distribution map



*Locations without exact coordinates are displayed as spherical country centroids

Accession number	UNITE taxon name	INSD taxon name	Sequence source	Interacting taxa	Sampling area	% ⓘ							← Alignment →		
						<.5	0.5	1.0	1.5	2.0	2.5	3.0			
JX434657	Tomentella	Tomentella (Tomentella sp. A11)													
UDB000238	Tomentella badia		TAAM159022 (Specimen)		Russian Federation										
HM370465	Thelephoraceae	Thelephoraceae (uncultured Tomentel...	Ectomycorrhiza		Italy										
**KM247748	Fungi	Fungi (uncultured fungus)			France										
**KJ188589	Thelephoraceae	Thelephoraceae (uncultured Telepho...		Neottia ovata	France										



Täna kuulamast!



Kessy Abarenkov



Allan Zirk



Plutof

