

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Lisett Klement
**TÜ Eesti geenivaramu geenidoonorite tervisega
seotud elukvaliteet**

Matemaatilise statistika eriala
Bakalaureusetöö (9 EAP)

Juhendajad: Prof. Krista Fischer, PhD
Taavi Lai, PhD

TARTU 2026

TÜ EESTI GEENIVARARMU GEENIDOOORITE TERVISEGA SEOTUD ELUKVALITEET

Bakalaureusetöö

Lisett Klement

Lühikokkuvõte

Inimese elukvaliteeti mõjutavad tegureid on palju ning üks olulisemaid on tervis. Selle mõju hindamine vaid diagnooside ning nende raskusastmete põhjal aga ei pruugi täpselt peegeldada seda, kuidas see indiviidi igapäevaelu mõjutab. Seetõttu on võetud kasutusele erinevaid instrumente, mille kaudu saavad inimesed hinnata oma tervisega seotud elukvaliteeti (*health-related quality of life*, HRQoL) ning üks sellistest on EuroQoli küsimustik.

Bakalaureusetöö eesmärk on hinnata Tartu Ülikooli Eesti geenivaramu aastatel 2009–2010 ja 2019–2020 liitunud geenidoonorite tervisega seotud elukvaliteeti EuroQoli küsimustiku abil, selle muutust kümne aasta jooksul ning seoseid erinevate kliiniliste ja demograafiliste teguritega.

Esmalt antakse ülevaade tervisega seotud elukvaliteedi olemusest ja EuroQoli küsimustikust ning tutvustatakse erinevaid meetodeid selle tulemuste kirjeldamiseks nagu LSS, HSDC ja Shannoni indeks. Lisaks käsitletakse ka K-keskmiste klasteranalüüsi ning dispersioonanalüüsi mudeli põhjal teostatavaid võrdluseid seoste uurimiseks. Töö teises pooles rakendatakse neid meetodeid ja tõlgendatakse tulemusi.

Selgub, et 2019.–2020. aastate geenidoonorid hindavad oma tervisega seotud elukvaliteeti paremini, kuid muutus sõltub tugevalt vanusest. Vanematel indiviididel on vähenenud probleemid kõigis dimensioonides peale valu, kus on hinnangud jäänud sarnaseks 2009.–2010. aasta seisuga. Noorematel aga on suurenenud probleemid nii igapäevategevustega toimetulekuga, valudega kui ka ärevusega, seejuures probleemid ärevusega on kasvanud enim. Lisaks täheldati, et üldjoontes hindavad enda tervisega seotud elukvaliteeti kõrgemalt nii kõrgharitud kui ka mehed.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika.

Märksõnad: EuroQol, tervisega seotud elukvaliteet, K-keskmiste klasteranalüüs.

HEALTH-RELATED QUALITY OF LIFE AMONG ESTONIAN BIOBANK GENE DONORS

Bachelor's thesis

Lisett Klement

Abstract

Many factors influence an individual's quality of life, one of the most significant being health. Assessing its impact solely on the basis of diagnoses and their severity may not accurately reflect how health affects an individual's daily life. Therefore, various instruments have been developed that allow individuals to assess their own health-related quality of life (HRQoL), one of which is the EuroQol questionnaire.

The aim of this bachelor's thesis is to evaluate the health-related quality of life of gene donors who joined the Estonian Biobank in 2009–2010 and 2019–2020, to analyse changes over a ten-year period and to examine associations with various clinical and demographic factors.

The first part provides an overview of the concept of health-related quality of life and the EuroQol questionnaire, followed by an introduction to different methods for its analysis, such as LSS, HSDC and the Shannon index. In addition, K-means clustering and analysis of variance are presented as methods for investigating relationships between different factors. In the second part, these methods are applied to the data and the results are interpreted.

The results indicate that later cohort of gene donors reported higher health-related quality of life, but the changes depended strongly on age. Among older individuals, problems have decreased in all dimensions except pain, where assessments have remained similar to the earlier period. Younger individuals have experienced an increase in problems related to usual activities, pain and anxiety, with anxiety having the biggest increase. It was also noted that men and more highly educated individuals tend to rate their health-related quality of life higher.

CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematics.

Key Words: EuroQol, health-related quality of life, K-means clustering.

Sisukord

Sissejuhatus	4
1 EuroQoli küsimustik EQ-5D-3L	5
2 Metoodika	7
2.1 Terviseseisundite tiheduskõver	7
2.2 Dimensioonide hinnangute summa	7
2.3 Shannoni indeks	7
2.4 K-keskmiste klasteranalüüs	9
2.4.1 Kүүnarnuki meetod klastrite arvu valimiseks	11
2.5 Dispersioonanalüüsi mudel	11
2.5.1 Keskmiste mitmene võrdlemine	13
3 Analüüs	15
3.1 Andmestiku kirjeldus	15
3.2 EuroQoli profiilid	16
3.2.1 Dimensioonide keskmiste hinnangute muutus vanusegruppides	21
3.2.2 K-keskmiste klasteranalüüs	25
3.3 EuroQoli indeksid	30
3.3.1 Võrdlus täiskasvanud elanikkonnaga	32
Arutelu	34
Kokkuvõte	37
Kasutatud allikad	39
Lisad	41

Sissejuhatus

Inimese elukvaliteeti mõjutavaid tegureid on palju – suhted, haridus, iseseisvus, tervis, kuuluvus, elukeskkond. Osa neist on kerge mõõta ja hinnata ning teisi keerukam. Indiviidi tervise mõju kirjeldatakse tihtipeale tema diagnooside ning nende raskusastmete kaudu, kuid see ei pruugi alati täpselt peegeldada terviseseisundi tegelikku mõju inimese elule. Selle jaoks, et täpsemini mõista, milline on see terviseseisund, mida indiviid oma elus kogeb, on loodud erinevaid tervisega seotud elukvaliteeti mõõtvaid instrumente. Nende kaudu on võimalik hinnata nii indiviidi kui ka rahvastiku tervisega seotud elukvaliteeti, mis muudab selle üheselt hinnatavaks ja võrreldavaks ning samuti on sisendiks kvaliteedi suhtes kohandatud eluaastate analüüsis, mille kaudu on võimalik luua otseseid seoseid inimese tervise, elukvaliteedi ning tervena elatud aastate vahel.

Üks tervisega seotud elukvaliteeti mõõtvatest instrumentidest on EuroQoli küsimustik ning see on osa Tartu Ülikooli Eesti geenivaramu (edaspidi ka geenivaramu) geenidonorite küsimustikust. Bakalaureusetöös hinnatakse geenivaramuga liitunud üldist tervisega seotud elukvaliteeti EuroQoli küsimustiku põhjal ja selle muutust kahe kohordi vahel. Võrreldavateks kohortideks on aastatel 2009–2010 ning 2019–2020 liitunud geenidonorid, kokku vastavalt 21 519 ja 76 728 indiviidi. Samuti uuritakse seoseid elukvaliteedi ning demograafiliste tegurite, nagu sugu, vanus ja haridus, kui ka Tervisekassa andmebaasist saadavate geenidonorite diagnooside vahel.

Töö esimeses pooles antakse esmalt ülevaade tervisega seotud elukvaliteedi mõõtmisest EuroQoli küsimustiku abil. Edasi tutvustatakse kasutatud meetodeid, mis sisaldavad erinevaid EuroQoli profiilide jaotust kirjeldavaid näitajaid, nagu LSS (*level sum score*), HSDC (*health state density curve*) ja Shannoni indeks, mille abil on võimalik saada üldine ülevaade mõlema kohordi elukvaliteedist ning kohortide iseärasustest. Lisaks käsitletakse K-keskmiste klasteranalüüsi ning keskmiste mitmest võrdlemist, mille abil uuritakse seoseid elukvaliteedi ja teiste tegurite vahel. Töö teises pooles rakendatakse antud meetodeid geenivaramu andmete peal ning tõlgendatakse tulemusi.

1 EuroQoli küsimustik EQ-5D-3L

EQ-5D-3L on lühike standardiseeritud küsimustik, mille abil on võimalik vastajal kirjeldada oma tervises seisundi mõju tema igapäevaelule, kasutades hinnanguid eluga rahulolu mõjutavate aspektide ehk dimensioonide kohta (Devlin, Parkin ja Janssen, 2020). EQ-5D-3L koosneb viiest dimensioonist, milleks on liikuvus, enesehooldus, igapäevaste tegevuste sooritamise võime, valu/ebamugavus ning ärevus/depressioon. Igal dimensioonil on kolm vastusevarianti – „pole probleeme“, „mõningad probleemid“, „äärmised probleemid“. (Lai, 2012)

EuroQoli küsimustik on üks viis hinnata tervisega seotud elukvaliteeti (*health-related quality of life*, HRQoL) (Devlin, Parkin ja Janssen, 2020). See hõlmab endas tervisega seotud tegurite kogumõju inimese heaolule ning selle alla ei kuulu näiteks majanduslikud ja poliitilised aspektid (Karimi ja Brazier, 2016). Arvestades küsimustiku üldisustaset, ei peegelda EQ-5D-3L kõigi tervisega seotud tegurite mõju, sest need sõltuvad olulisel määral ka konkreetsetest diagnoosidest (Devlin, Parkin ja Janssen, 2020).

EuroQoli küsimustiku tulemusi saab analüüsida kahe erineva näitaja põhjal, milleks on profiilid ja indeksid. Profiil võtab lühidalt kokku küsimustikust saadud hinnangud viie dimensiooni kohta. Kõige kompaktsem viis selle koostamiseks on teisendada indiviidi vastused viienumbriks sõnaks, kus järjekorranumber määrab eeltoodud järjestuse põhjal vastava dimensiooni ning selle väärtus vastab valitud vastusevariandile. Näiteks „13212“ tähendab, et vastajal ei ole probleeme liikumisega ega valu/ebamugavusega, samas on suured probleemid enese eest hoolitsemisega ning mõningad probleemid igapäevaste tegevuste sooritamise ja ärevuse/depressiooniga. Kokku on seega $3^5 = 243$ võimalikku profiili. Indeksid võtavad ühe arvulise väärtusena kokku profiili üldseisundi, mis muudab tulemused lihtsamini võrreldavaks ja tõlgendatavaks ning need on ka sisendiks kvaliteedi suhtes kohandatud eluaastate (*quality-adjusted life years*, QALYs) analüüsis. Indeksite arvutamiseks väärtustatakse iga dimensiooni hinnangut vastavalt sellele, kui suur on selle mõju üldpopulatsioonis. Näiteks probleemid valu ja ebamugavusega on sageli suurema kaaluga kui mured enesehooldusega, kuid kaalude väärtused sõltuvad samuti demograafilistest

ja geograafilistest teguritest. Varasemalt vaadeldud profiilile „13212“ vastava indeksi väärtuse leidmiseks lisas 6 toodud Euroopa *value set*’i põhjal, arvutatakse esmalt karistustegur:

$$\begin{aligned} & (\text{Liikuvus} = 1) + (\text{Enesehoolitsus} = 3) + (\text{Igapäevatoimetused} = 2) + \\ & (\text{Valu} = 1) + (\text{Ärevus} = 2) + (\text{Suurem kui 1}) + (\text{Suurem kui 2}) = \\ & 0 + 0,1559 + 0,0264 + 0 + 0,0891 + 0,2288 + 0,1279 = 0,6281 \end{aligned}$$

Indeksi väärtuse saame, kui lahutame selle ühest ehk $1 - 0,6281 = 0,3719$. Selle arvutamisel tuleb valida lähtuvalt valimist sellele sobiv väärtuste kogum (varasemalt mainitud *value set*). Iga kogumi korral on indeksi maksimumväärtus 1, mis vastab profiilile „11111“ ja esindab parimat võimalikku elukvaliteeti, ning miinimumväärtus sõltub valitud kogumist. Kogumite konstrueerimine on keeruline ja aeganõudev protsess, mistõttu paljudel riikidel puudub nende rahvastiku jaoks kalkuleeritud spetsiaalne kogum. Sellises olukorras on EuroQoli küsimustiku koostajate soovitus kasutada geograafiliselt lähima riigi kogumit, kui see leidub. Samas võib olla põhjendatud analüüsida ka teiste riikide kogumeid, eriti kui valimi eripära seda toetab, võrrelda saadud tulemusi ning selle põhjal otsustada, milline kaalude kogum esindab valimit kõige paremini. (Devlin, Parkin ja Janssen, 2020) Näiteks Šveitsi vähihaigete EuroQoli kogumi valimiseks võrreldi nii Saksamaa, Prantsusmaa, Ühendkuningriigi kui ka Euroopa variante, mis olid valitud lähtuvalt geograafilisest lähedusest või Šveitsis elavate mitte-šveitslaste osakaalust valimis (Matter-Walstra *et al.*, 2014).

2 Metoodika

2.1 Terviseseisundite tiheduskõver

Selle jaoks, et hinnata, kuidas on jaotunud profiilid valimis ning kui tugev on nn „lae efekt“ andmestikus, on otstarbekas kasutada terviseseisundite tiheduskõverat (*health state density curve*, HSDC) (Devlin, Parkin ja Janssen, 2020). „Lae efektiks“ nimetatakse olukorda, kus väga hea terviseseisundiga inividid moodustavad nii suure osa valimis, et erinevused väga hea tervisega inimeste vastuste vahel ei avaldu piisavalt hästi (Haqiqatkah, 2025). See on EuroQoli tulemuste puhul tavaline nähtus, mistõttu on oluline hinnata selle tugevust. Meetodi eesmärk seisneb profiilide kumulatiivsete sageduste võrdlemises valimi kumulatiivsete sagedustega. Kui profiilid on valimis ühtlastelt jaotunud, on tulemiks 45° tõusuga sirge, mis tähendab, et $x\%$ vaatlustest katab täpselt $x\%$ profiilidest. Mida kaugemal on saadud kõver sellest sirgest, seda rohkem kuhjunud on profiilid. (Devlin, Parkin ja Janssen, 2020) Meetodi visuaalne rakendus on leitav analüüsi osast jooniselt 2.

2.2 Dimensioonide hinnangute summa

Dimensioonide hinnangute jaotuse uurimiseks kasutatakse hinnangute summat (*level sum score*, LSS). Selle arvutamiseks liidetakse kokku kõigi dimensioonide hinnangud ning EQ-5D-3L korral jääb selle väärtus vahemikku 5–15, kus väärtus 5 vastab profiilile „11111“ ning 15 profiilile „33333“. Siinkohal tuleb tähele panna, et see näitaja ei arvesta, mis dimensioonis mõni hinnang on antud – nii „11131“ kui ka „13111“ korral on LSS väärtuseks $1 + 1 + 1 + 1 + 3 = 1 + 3 + 1 + 1 + 1 = 7$, kuid suured probleemid valu või ebamugavusega on tavaliselt märkimisväärsema mõjuga elukvaliteedile kui probleemid enesehoolitusega. (Devlin, Parkin ja Janssen, 2020)

2.3 Shannoni indeks

Olgu meil diskreetne juhuslik suurus X jaotusega P . Olgu $\mathcal{X} = \{x_1, x_2, \dots\}$ üli-
malt loenduv hulk, mis sisaldab juhusliku suuruse X kõiki võimalikke väärtusi, ning

tõenäosus, et X võtab väärtuse x_i , defineeritud kui

$$p_i := P(X = x_i) = P(x_i).$$

Juhusliku suuruse X (Shannoni) entroopia $H(X)$ on

$$H(X) = - \sum_i p_i \log_2 p_i = - \sum_{x \in \mathcal{X}} P(x) \log_2 P(x) = - \sum_{x \in \mathcal{X}_P} P(x) \log_2 P(x).$$

Entroopia mõõdab juhusliku suuruse X keskmist juhuslikkust – suurema juhuslikkuse korral on selle väärtus raskemini ette ennustatav ning seda rohkem informatsiooni selle teada saamine annab. (Lember, 2022)

Kui Shannoni entroopia abil kirjeldatakse mingi liigi mitmekesisust, kasutatakse tihti selle asemel nimetust Shannoni indeks. EuroQoli küsimustiku korral saab selle indeksi abil hinnata profiilide või dimensioonide tasemete informatiivsust ehk mitmekülgust. (Devlin, Parkin ja Janssen, 2020) See on defineeritud kui

$$H' = - \sum_{i=1}^C p_i \log_2 p_i,$$

kus C on võimalike kategooriate arv (dimensioonitasemed või profiilid) ja $p_i = \frac{n_i}{N}$, kus n_i on i -ndasse kategooriasse kuulunud vaatluse arv ning N on valimi kogumaht. Kõrgem indeksi väärtus viitab vastuste suuremale mitmekesisusele ning ühtlasemale jaotusele kategooriate vahel ning väiksem väärtus kuhjumisele väiksemale osale kõigest võimalikest valikutest. Indeks saavutab maksimaalse väärtuse $H'_{max} = \log_2 C$, kui kõik kategooriad on võrdselt tõenäolised.

Shannoni indeksi dispersioon on hinnatav kui

$$Var(H') = \frac{\sum_{i=1}^C p_i (\log_2(p_i))^2 - (\sum_{i=1}^C p_i \log_2(p_i))^2}{N},$$

mille abil on võimalik arvutada standardviga $SE = \sqrt{Var(H')}$ ning 95%-usaldusintervall $(H' - 1,96 \cdot SE, H' + 1,96 \cdot SE)$. (Janssen, Birnie ja Bonsel, 2007)

2.4 K-keskmiste klasteranalüüs

Järgnev peatükk põhineb teosel „An introduction to statistical learning with applications in R“ (James *et al.*, 2023), kui ei ole märgitud teisiti.

K-keskmiste klasteranalüüs on juhendamata õppe meetod, mille eesmärgiks on jagada andmestik K -sse mittekattuvasse klastrisse, kus K on ette antud parameeter, mis tähistab klastrite arvu.

Olgu meil n vaatlust, millel on mõõdetud p tunnust. Tähistagu x_{ij} i -nda vaatluse j -nda tunnuse väärtust ja olgu $C_1, \dots, C_K$ hulgad (klastrid), mis sisaldavad nendes kuuluvate vaatluste indekseid ning mille korral kehtib

1. $C_1 \cup C_2 \cup \dots \cup C_K = \{1, \dots, n\}$ ehk iga vaatlus kuulub vähemalt ühte klastrisse,
2. Kui $k \neq k'$, siis $C_k \cap C_{k'} = \emptyset$ ehk ükski vaatlus ei saa kuuluda rohkem kui ühte klastrisse.

Klastrid C_k klasterisisene hajuvus $W(C_k)$ mõõdab, kui palju sellesse klastrisse kuuluvad vaatlused teineteisest erinevad. Seda saab defineerida eukleidilise kauguse ruudu abil. Kahe p -mõõtmelise vektori x_i ja $x_{i'}$ vaheline eukleidiline kaugus on defineeritud kui

$$d(x_i, x_{i'}) = \sqrt{(x_{i1} - x_{i'1})^2 + (x_{i2} - x_{i'2})^2 + \dots + (x_{ip} - x_{i'p})^2} = \sqrt{\sum_{j=1}^p (x_{ij} - x_{i'j})^2}$$

ning seega klasterisisene hajuvus $W(C_k)$ avaldub kujul

$$W(C_k) = \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2,$$

kus $|C_k|$ on k -ndasse klastrisse kuuluvate vaatluste arv. Klasteranalüüsi eesmärgiks on minimeerida klasteritesest hajuvust ehk

$$\min_{C_1, \dots, C_K} \left\{ \sum_{k=1}^K W(C_k) \right\} = \min_{C_1, \dots, C_K} \left\{ \sum_{k=1}^K \frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right\}. \quad (1)$$

K-keskmiste klasteranalüüsi algoritme on erinevaid, kuid antud töös on kasutatud Hartigani ja Wongi algoritmi (Hartigan ja Wong, 1979), mis on rakendustarkvaras R vaikimisi kasutatav algoritm. Üldine algoritmide loogika on järgmine.

1. Määrame esialgsed klatrikeskmised ehk tsentroidid, mis avaduvad p -mõõtmeteliste vektoritena $m_k = (m_{k1}, \dots, m_{kp})$, $k = 1, \dots, K$. Tsentroidide valimiseks on erinevaid võimalusi – jagada kõik vaatlused suvaliselt klatriksesse ning arvutada nende klatriksite keskmised või valida välja K suvalist vaatlust ning kasutada neid esialgsete klatriksite tsentroididena. Antud töös kasutatakse teist varianti ehk esialgsed klatrikskeskmised on suvalised K vaatlust.
2. Kordame järgmist algoritmi kuni vaatlused jäävad püsivalt ühte klatriksisse.
 - (a) Määrame iga vaatluse klatriksisse, mille korral antud vaatluse eukleidiline kaugus klatrikskeskmisest on vähim.
 - (b) Arvutame iga klatri C_k , $k \in \{1, \dots, K\}$, jaoks uue klatrikskeskmise m_k , mille j -ndaks elemendiks on vastavasse klatriksisse kuuluvate vaatluste j -nda tunnuse keskmine väärtus ehk

$$m_{kj} = \frac{1}{|C_k|} \sum_{i \in C_k} x_{ij}.$$

Selle jaoks, et mõista, miks sammud 2.(a) ja 2.(b) vähendavad klatriksitesisest hajuvust, vaatleme järgmist võrdust:

$$\frac{1}{|C_k|} \sum_{i, i' \in C_k} \sum_{j=1}^p (x_{ij} - x_{i'j})^2 = 2 \sum_{i \in C_k} \sum_{j=1}^p (x_{ij} - m_{kj})^2. \quad (2)$$

See tähendab, et klatriksisse C_k kuuluvate vaatluste omavaheliste kauguste ruutude keskmine väärtus on võrdne kahekordse kõigi vaatluste kauguste ruutude summa klatriks keskmisest m_{kj} . Meie eesmärk oli minimeerida avaldist (1), seega minimeerides kõigi vaatluste kauguste ruutude summat klatriks keskmisest, minimeerime võrduse (2) põhjal ka klatriksisse kuuluvate vaatluste keskmist omavahelist kauguste ruutude summat, sest need kaks avaldist erinevad vaid konstandi poolest. Seejuures

saab vaatluste ümberpaigutamine uutesse klastritesse ainult parandada antud tulemust, mistõttu ei saa iteratsioonide käigus kunagi klastritesisene hajuvus suurenedagi. Iteratsioone tehakse nii kaua kuni vaatluste jaotus klastrate vahel enam ei muutu – on leitud lokaalne miinimum. Kuna algoritm saavutab lokaalse, kuid mitte globaalse, miinimumi, sõltub tulemus otseselt algoritmi alguses määratud klatrikeskmistest. Seetõttu on oluline algoritmi rakendada mitu korda ning valida saadud tulemustest parim näiteks klastritesisese hajuvuse põhjal.

2.4.1 Kūūnarnuki meetod klastrate arvu valimiseks

K-keskmiste klasteranalüüsi tulemus sõltub lisaks esialgsete tsentroidide valimisele ka klastrate arvust. On olemas erinevaid meetodeid optimaalse klastrate arvu määramiseks, et tuvastada andmestikus sisulisi alarühmi ning vältida müra klasterdamist.

Kūūnarnuki meetod (*elbow method*) on visuaalne hindamismeetod, mida kasutatakse selleks, et valida optimaalne klastrate arv. Idee seisneb selles, et klasterdamist viiakse läbi mitu korda erinevate klatriarvude korral, alustades $K = 2$ klatriist ja suurendades seda iga järgmine kord ühe võrra, ning hinnatakse jooniselt, millise klatriarvu korral ei toimu selle suurendamisel enam olulist vähenemist klatriisisese hajuvuses. Optimaalne klastrate arv on punkt alates millest on visuaalselt näha, et klatriisisese hajuvuse vähenemine muutub märgatavalt aeglasemaks. (Kodinariya ja Makwana, 2013)

2.5 Dispersioonanalüüsi mudel

Dispersioonanalüüsi mudelit kasutatakse selleks, et hinnata faktori(te) mõju uuritava tunnusele, kasutades selleks rühmade keskväärtuste võrdlemist. Kontrollitavad hüpoteesid on

$$H_0 : \mu_1 = \dots = \mu_m$$

$$H_1 : \exists i, j : \mu_i \neq \mu_j$$

(Käärik, [s.a.](#))

Üldiseid lineaarseid mudeleid saab esitada maatrikskujul:

$$y = X\beta + \epsilon. \quad (3)$$

Kui valimimaht on n ning mudelis on p hinnatavat parameetrit, siis

- y – $n \times 1$ vektor, mis sisaldab uuritava tunnuse väärtusi,
- X – $n \times p$ plaanimaatriks, kus iga rida vastab ühele vaatlusele ning veerud sisaldavad seletavate tunnuste väärtusi,
- β – $p \times 1$ vektor, mis koosneb tundmatutest parameetritest, mida soovitakse hinnata,
- ϵ – $n \times 1$ juhuslike vigade vektor.

(Sarty, [2022](#))

Järgnevalt tutvustame seda maatrikskujut täpsemalt juhul, kui argumenttunnused on kategoriseerivad. Sellisel juhul on tegemist dispersioonanalüüsi mudeliga.

Näiteks vaatleme kolmetasemelist tunnust, mille iga tase sisaldab kahte vaatlust.

Olgu $y_i(j)$ i -ndasse tasemesse kuuluv j -s vaatlus. Sellisel juhul oleks mudel kujul

$$y = \begin{bmatrix} y_1(1) \\ y_1(2) \\ y_2(1) \\ y_2(2) \\ y_3(1) \\ y_3(2) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1(1) \\ \epsilon_1(2) \\ \epsilon_2(1) \\ \epsilon_2(2) \\ \epsilon_3(1) \\ \epsilon_3(2) \end{bmatrix} = X\beta + \epsilon.$$

Plaanimaatriksi X esimene veerg on vabaliikme veerg ning ülejäänud $m - 1$ veergu, kui m on tasemete arv, vastavad vaadeldud tunnuse tasemetele võrreldes referentskategoriaga, mis antud näites on vahepeal esimene tase. (Sarty, [2022](#))

Ülaltoodud viisil defineeritud mudeli puhul kehtib $\beta_1 = \mu_1$ ning $\beta_i = \mu_i - \mu_1$, $i = 2, \dots, m$. Seega kehtib H_0 , kui $\beta_2 = \dots = \beta_m = 0$, sest siis iga i korral $\mu_i - \mu_1 = 0$ ehk $\mu_i = \mu_1$.

Tundmatud parameetrid hinnatakse vähimruutude meetodil võrrandisüsteemi $(X^T X)\beta = X^T y$ lahenditena ning parameetrite hinnangud avalduvad vektorina $\hat{\beta} = (X^T X)^{-1} X^T y$. (Sarty, 2022)

Mudeli klassikalisteks eeldusteks on

1. mudeli vead on normaalse jaotusega,
2. sõltuva tunnuse hajuvus on faktortunnuse tasemetel ühesugune (dispersiooni-de homogeensus),
3. mõõtmised on sõltumatud.

(Käärik, [s.a.](#))

2.5.1 Keskmiste mitmene võrdlemine

Kui dispersioonanalüüsi tulemusena on teada, et faktori(tel) on mõju, on otstarbekas uurida milliste faktorite tasemed teineteisest erinevad. Selleks on mitmeid variante, kuid antud tööd kasutatakse vähimruutude keskmiste t-teste ning rakendatakse Bonferroni korrigeerimist, et hoida kontrolli all I liiki vea tegemise tõenäosust.

Vähimruutude keskmised on mudelist hinnatud mõjude lineaarkombinatsioonid, mis baseeruvad kasutataval mudelil. (Käärik, [s.a.](#))

Võrrandis (3) oleval vigade vektoril ϵ on $n - p$ vabadusastet. Vigade dispersiooni saab leida valemiga

$$\sigma^2 = \frac{\epsilon^T \epsilon}{n - p},$$

mille abil saab kahe keskmise võrdlemiseks leida statistiku kujul

$$t = \frac{c^T (\hat{\beta} - \beta_0)}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}}, \quad (4)$$

kus tavaliselt $\beta_0 = \vec{0}$ ning c on $p \times 1$ vektor, mis määrab, milliste gruppide keskmisi omavahel võrrelda. Näiteks teise ja kolmanda grupi võrdlemisel on vektor c kujul

$$c^T = \begin{pmatrix} 0 & 1 & -1 & 0 & \dots & 0 \end{pmatrix}$$

(Sarty, 2022)

Kui $L = c^T \beta$, siis t-testiga kontrollitakse hüpoteese

$$H_0 : L = 0$$

$$H_1 : L \neq 0$$

Kui H_0 kehtib, siis on statistik (4) t-jaotusega vabadustastmetega $n - p$. (Käärrik, *s.a.*)

Keskmiste võrdlemine on seotud kahte erinevat tüüpi vea tegemisega:

1. Võrdlusviisilise vea korral soovitakse kontrollida I liiki vea tegemiste tõenäosust ühe konkreetse võrdluspaari korral,
2. Katseviisilise vea korral soovitakse kontrollida vähemalt ühe I liiki vea tegemise tõenäosust kõigi võrreldavate paaride seas.

(Käärrik, *s.a.*)

Selle jaoks, et hoida mitmese võrdlemise korral katseviisilist viga kontrolli all, saab kasutada Bonferroni korrigeerimist. Kui soovime, et katseviisilise vea tõenäosus oleks väiksem või võrdne α -ga, tuleb võtta võrdlusviisilise vea tõenäosuseks $\frac{\alpha}{m}$, sest selisel juhul tõenäosus, et mitte ühegi võrdluse korral ei tehta viga lubatust suurema tõenäosusega, on $(1 - \frac{\alpha}{m})^m \approx 1 - \alpha$. (Bland ja Altman, 1995)

3 Analüüs

3.1 Andmestiku kirjeldus

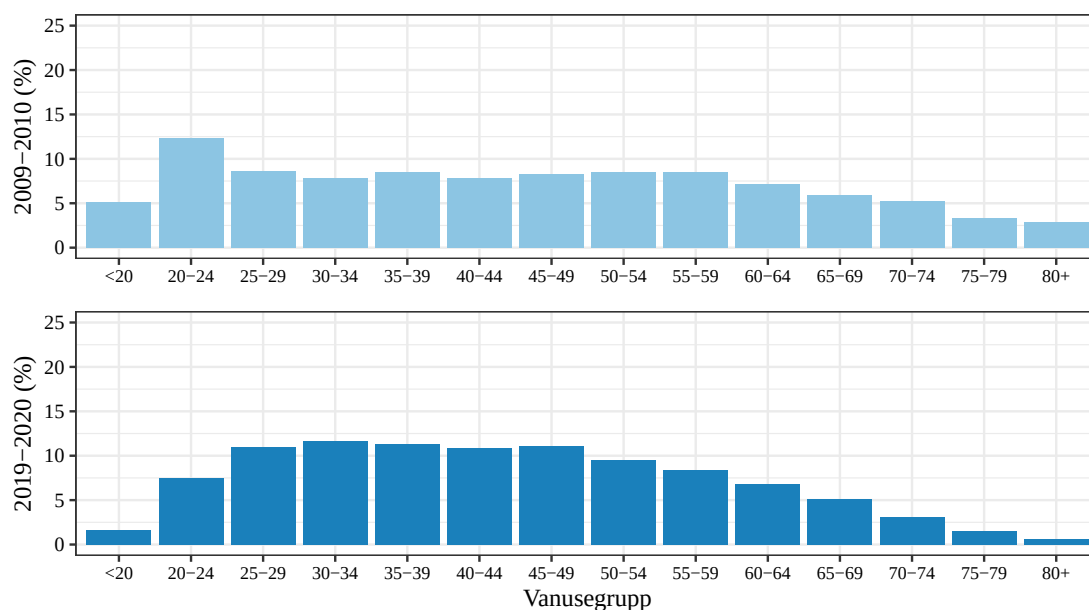
Töö praktiline osa põhineb Tartu Ülikooli Eesti geenivaramu geenidoonorite 2009.–2010. ja 2019.–2020. aastate andmetel. Esimese perioodi valimisse kuulus 21 519 inimest ning teisse 76 728 inimest.

Esimese kohordi keskmine vanus oli 45,16 aastat ning sellest 62,80% moodustasid naised ja 37,20% mehed. Teise kohordi keskmine vanus oli väiksem, 43,86 aastat, ning naiste osakaal vähesel määral suurem (68,54% naisi ja 31,46% mehi). Samuti oli teises kohordis rohkem kõrgharitud (68,29%) ning vähem põhihariduse või madalama haridusastmega doonoreid (2,89%). (tabel 1)

Tabel 1: Tunnuste ülevaade

Tunnus	2009–2010	2019–2020
Doonorite arv	21519	76728
Mehed	8006 (37,20%)	24140 (31,46%)
Naised	13513 (62,80%)	52588 (68,54%)
Keskmine vanus	45,16	43,86
Haridusaste		
Põhiharidus	2320 (10,78%)	2216 (2,89%)
Keskharidus	7392 (34,35%)	22115 (28,82%)
Kõrgharidus	11807 (54,87%)	52397 (68,29%)

Aastatel 2009–2010 olid ülekaalukamalt esindatud nii nooremad kui ka vanemad vanusegrupid. Alla 24-aastaste osakaal oli 17,4% ja üle 70-aastaste 11,41%, samas kui vastavad osakaalud aastatel 2019–2020 olid 9,03% ning 5,3%. Varasem kohort moodustus suures osas 25–64-aastastest, kus iga 5-aastane vanusegrupp sellest vahemikust moodustas ligikaudu 7–8% kogu valimist. 2019.–2020. aastate kohordis moodustas valdava enamuse aga 25–49-aastased ning selle vahemiku 5-aastaste vanusegruppide osakaalud jäid 11–12% piiridesse. (joonis 1)



Joonis 1: Vanusegruppide jaotus kohortides

On oluline märkida, et vaadeldava kahe kohordi kaasamismeetodid ei olnud samad ning sellel võib olla mõju kohorti kuulunud inimeste tervisele. Varasemal perioodil kaasati doonoreid põhiliselt terviseasutustest (apteegid, perearstid), mistõttu moodustub see kohort põhiliselt indiviididest, kes mingi tervisemure tõttu sellistesse asutustesse pöörduma pidid. Aastatel 2019–2020 aga oli geenidoonorlus kättesaadavam laiemale rahvastikule.

3.2 EuroQoli profiilid

Järgnevalt analüüsitakse EuroQoli dimensioonide hinnanguid ning nendest moodustuvaid profile. EuroQoli dimensioonide skaalal tähistasid suuremad väärtused suuremaid probleeme, samas kui väiksemad väärtused viitasid vähesematele probleemidele.

2009.–2010. aastate valimis oli esindatud 116 ning aastatel 2019–2020 132 erinevat profiili 243st võimalikust. Esimese kohordi kõige sagedasemini esinev profiil oli „11111“ ehk ideaalne tervis ning see moodustas 43,64% valimist (tabel 2). Seejuures oli 95,34% valimist kirjeldatav profiilide kaudu, milles ükski dimensioon ei sisaldanud

hinnangut „3“. Ka teises kohordis oli kõige sagedasem profiil „11111“, moodustades ligikaudu sama suure osa kogu valimist (43,31%) (tabel 3). Võrreldes esimese kohordiga oli aga 96,57% profiilidest teises kohordis ilma ühegi dimensiooni hinnanguta, mille väärtus oleks „3“ ehk ekstreemsete probleemidega inimeste osakaal oli 1,23 protsendipunkti võrra väiksem. Kümnest sagedasemast profiilist mõlemas kohordis kattuvad 9 ning aastatega on muutunud vaid nende sagedused. Profiil, mis kuulus sagedasemate hulka aastatel 2009–2010, aga ei esinenud nii suurel määral aastatel 2019–2020, oli „22222“, mis tähendab mõningaid probleeme kõigis dimensioonides. Selle asemel oli hilisemas kohordis enim esinenud profiilide hulgas „21121“ ehk mõningad probleemid valu ja ebamugavuse ning liikuvusega.

Tabel 2: 2009.–2010. aastate kohordi 10 sagedasemat profiili

Profiil	n (% valimist)	Kumulatiivne %
11111	9391 (43,64%)	43,64%
11112	2479 (11,52%)	55,16%
11122	2346 (10,90%)	66,06%
11121	1678 (7,80%)	73,86%
21222	942 (4,38%)	78,24%
11222	820 (3,81%)	82,05%
22222	507 (2,36%)	84,41%
21221	500 (2,32%)	86,73%
11221	350 (1,63%)	88,36%
21122	303 (1,41%)	89,77%

Tabel 3: 2019.–2020. aastate kohordi 10 sagedasemat profiili

Profiil	n (% valimist)	Kumulatiivne %
11111	33234 (43,31%)	43,31%
11112	11029 (14,37%)	57,68%
11121	9405 (12,26%)	69,94%
11122	8617 (11,23%)	81,17%
21222	2017 (2,63%)	83,80%
11222	1764 (2,30%)	86,10%
21221	1592 (2,07%)	88,17%
21121	1447 (1,89%)	90,06%
11221	1125 (1,47%)	91,53%
21122	1089 (1,42%)	92,95%

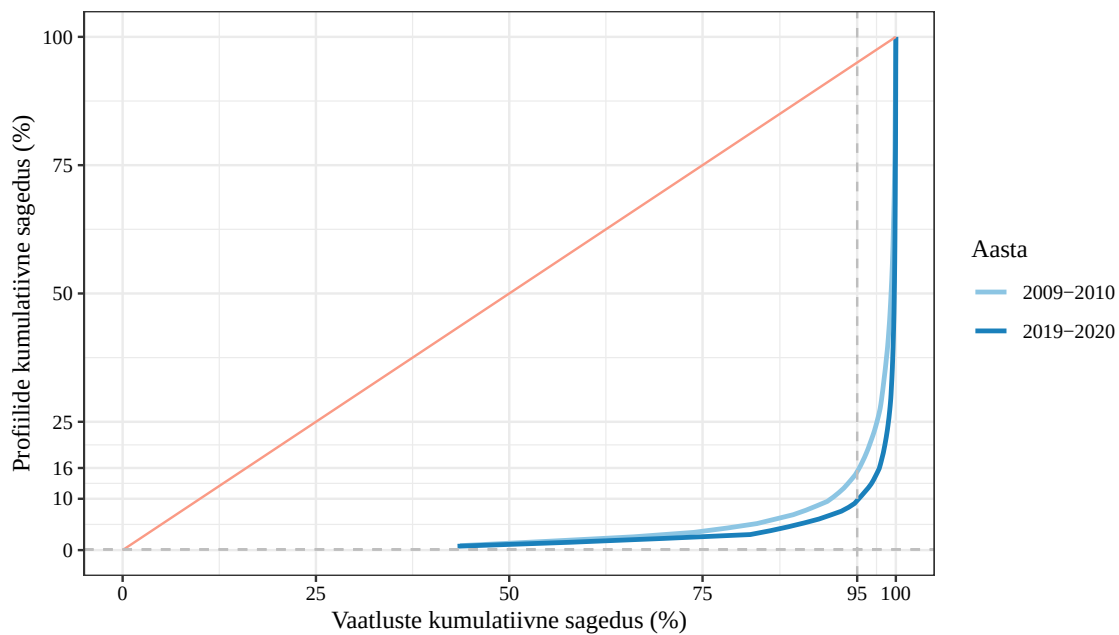
Ideaalse tervisehinnanguga inimeste osakaal oli mõlemas kohordis suurem mees-
tel. 2009.–2010. aastatel esines naiste seas meestest rohkem profiili, mis kujutas

mõningaid probleeme ärevuse ja depressiooni ning valu ja ebamugavusega (profiil „11122“) ning ilmnis rohkem ka neid, kellel esines mõningaid probleeme kõigis dimensioonides peale enesehoolitsuse (profiil „21222“). Aastatel 2019–2020 oli protsentuaalselt suurem profiilide „11112“ ning „11121“ osakaal, kuid eelnevalt vaadeldud „11122“ osakaal jäi ligikaudu samaks. Ka meeste seas oli muutunud profiil „11121“ sagedasemaks. (tabel 4)

Tabel 4: Kohortide 5 sagedasemat profiili sugude lõikes ning nende esinemiste sagedused

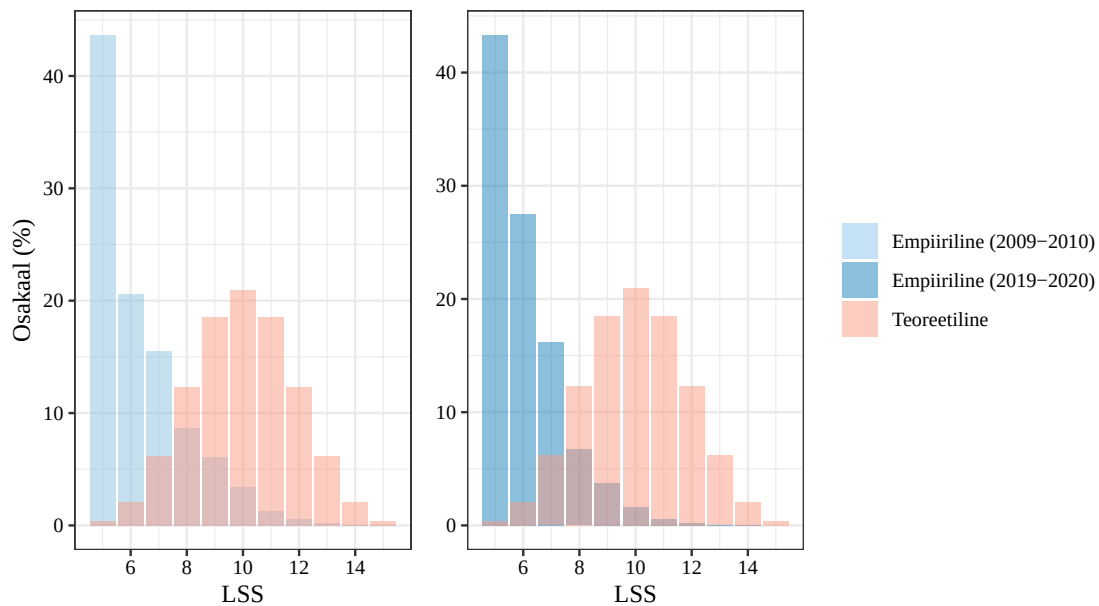
Profiil	Mehed	Profiil	Naised
2009–2010			
11111	3704 (46,27%)	11111	5687 (42,09%)
11112	897 (11,20%)	11122	1600 (11,84%)
11122	746 (9,32%)	11112	1582 (11,71%)
11121	653 (8,16%)	11121	1025 (7,59%)
11222	280 (3,50%)	21222	664 (4,91%)
2019–2020			
11111	11279 (46,72%)	11111	21955 (41,75%)
11112	3255 (13,48%)	11112	7774 (14,78%)
11121	3011 (12,47%)	11121	6394 (12,16%)
11122	2352 (9,74%)	11122	6265 (11,91%)
21222	530 (2,20%)	21222	1487 (2,83%)

Jooniselt 2 on näha, et 2009.–2010. aastatel kattis 95% vaatlustest 16% valimis esinenud profiilidest, kuid 2019.–2020. aastatel 10% – teisel vaatlusperioodil olid esinenud profiilid rohkem kuhjunud ning väiksem osa profiile kirjeldas suuremat osa valimist, kuid see võib olla tingitud ka kohortide valimimahtude erinevusest. Samuti ilmneb, et laaeft esines mõlemas kohordis, kuna vastused koonduvad skaala ülemisse ossa, kuid see efekt on kohortide vahel väga sarnane, millele viitab ka eelnevalt täheldatud asjaolu, et mõlemas kohordis oli profiili „11111“ osakaal sarnane (43,64% ja 43,31%).



Joonis 2: Kohortide HSDC kõverad

Uurides 2009–2010 ja 2019–2020 kohortide LSSi jaotust võrdsete profilitõenäosustega teoreetilise jaotusega, kus iga profiili esines vaid ühe korra, ilmneb, et empiiriline jaotus oli nihkes madalamate LSSi väärtuste poole, st geenidoonoritel esines rohkem parema elukvaliteediga profile. Täpsemalt olid suurema osakaaluga profiilid, mille LSS jääb vahemikku 5–7. Võrreldes 2009.–2010. aastate kohordiga olid hilisemas kohordis märgatavalt suurema osakaaluga profiilid, mille LSS vastab väärtusele 6, kuid selle võrra oli esimeses kohordis rohkem esindatud profiilid LSS väärtustega vahemikust 8–11. See viitab asjaolule, et esimeses kohordis esines rohkem halvema elukvaliteediga profile ehk neid, kelle puhul üks või mitu dimensiooni oli hinnanguga „3“ või rohkem kui üks dimensioon hinnanguga „2“. (joonis 3)



Joonis 3: LSSi väärtuste empiirilised ja teoreetiline jaotus

Shannoni indekseid põhjal selgub, et valu ja ebamugavuse (edaspidi valu) ning ärevuse ja depressiooni (edaspidi ärevuse) dimensioonides ei ole toimunud märkimisväärt muutust hinnangute varieeruvuses. Liikuvuse, enesehoolitsuse ja igapäevategevustega toimetuleku (edaspidi igapäevatoimetuste) hinnangute Shannoni indekseid 95%-usaldusintervallid ei kattu, mis tähendab, et aastatel 2019–2020 oli vähenenud hinnangute mitmekülgsus nendes dimensioonides ehk hinnangud olid koondunud rohkem väiksemale osale kõigist võimalikest variantidest. (tabel 5) Liikuvuse dimensioonis oli hinnangu „1“ osakaal kasvanud 5,93 PP, enesehoolitsuses 3,69 PP ja igapäevatoimetustes 8,01 PP võrra. Samas oli valu dimensioonis selle hinnangu osakaal tõusnud vaid 1,62 PP ning ärevuse oma 2,47 PP võrra. Ärevuse korral olid kasvanud äärmiste hinnangute osakaalud, st vähenenud oli hinnangu „2“ osakaal, kuid selle arvelt olid tõusnud nii hinnangu „1“ kui ka hinnangu „3“ osakaalud, millest viimane tõusis 0,36 protsendipunkti. Teistes dimensioonides oli kasvanud vaid hinnangu „1“ osakaal. (vt. täpsemalt lisast 1) Seega oli 2009–2010 perioodi geenidoonoritel rohkem probleeme liikuvuse, enesehoolituse ja igapäevaste toimingutega ning 2019.–2020. aastatel olid need probleemid vähenenud, aga valu ja ärevuse hinnangud olid jäänud sarnasemaks või halvenenud võrreldes varasema

perioodi kohordiga.

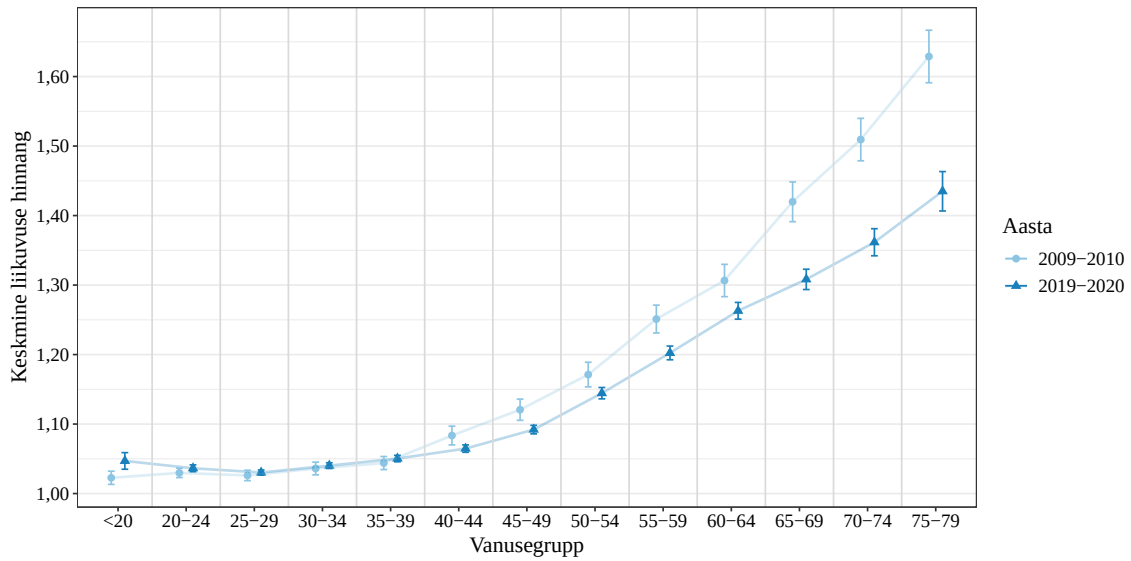
Tabel 5: Dimensioonide hinnangute Shannoni indeksid

Dimensioon	Shannoni indeks (95%-usaldusvahemik)			
	2009–2010		2019–2020	
Liikuvus	0,708	(0,695, 0,721)	0,533	(0,526, 0,540)
Enesehoolitsus	0,366	(0,352, 0,380)	0,181	(0,175, 0,187)
Igapäevased toimingud	0,859	(0,844, 0,873)	0,603	(0,595, 0,610)
Valu	1,064	(1,055, 1,073)	1,055	(1,050, 1,059)
Ärevus	1,071	(1,061, 1,080)	1,067	(1,062, 1,073)

3.2.1 Dimensioonide keskmiste hinnangute muutus vanusegruppides

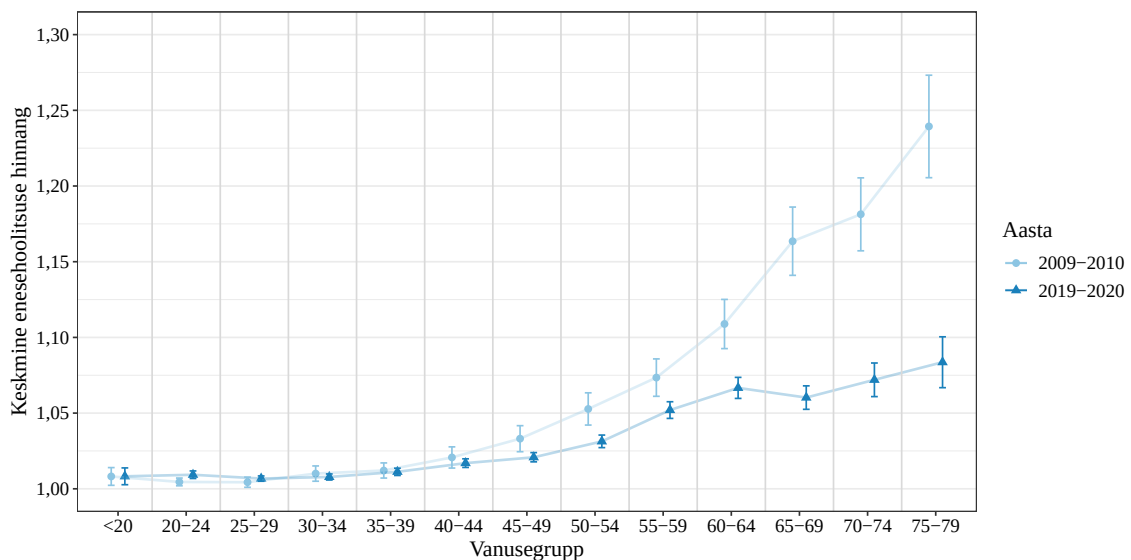
Järgnevas peatükis uurime põhjalikumalt, kuidas muutuvad hinnangud erinevates dimensioonides vanusegruppide ja kohortide lõikes. Lisaks usaldusvahemike võrdlemisele viidi iga dimensiooni korral läbi ka vähimruutude keskmiste mitmene võrdlemine koos Bonferroni korrektsiooniga. Mitmese võrdlemise tulemused koos keskmiste täpsete väärtustega on leitavad lisast 5, seejuures jäeti analüüsist välja 80-aastased ja vanemad, kuna nende osakaal oli väike ning jaotus kohortide vahel ebaproportsionaalne. Samuti ei ole see vanusegrupp ülemise piiriga fikseeritud, mistõttu võib selle sisemine vanuseline koosseis kohortide lõikes erineda.

Liikuvuse keskmised hinnangud jäid vahemikku 1,01 – 1,67 ehk varieerusid kuni 0,66 ühikut. Kohortide vahel ei erinenud hinnangud olulisel määral 40–44-aastaste vanusegruppi ning peale seda olid 2019.–2020. aastate geenidoonorite keskmised hinnangud kõigis vanusegruppides oluliselt madalamad, seejuures vanusegruppi kasvades suurenes ka keskmiste hinnangute erinevus kohortide vahel. Seega oli kümme aastat hilisemas kohordis vähem probleeme liikuvusega vanematel geenidoonoritel, nooremate seas on aga keskmise hinnangu muutumatus ootuspärane, sest keskmine oli juba aastatel 2009–2010 väga madal. (joonis 4)



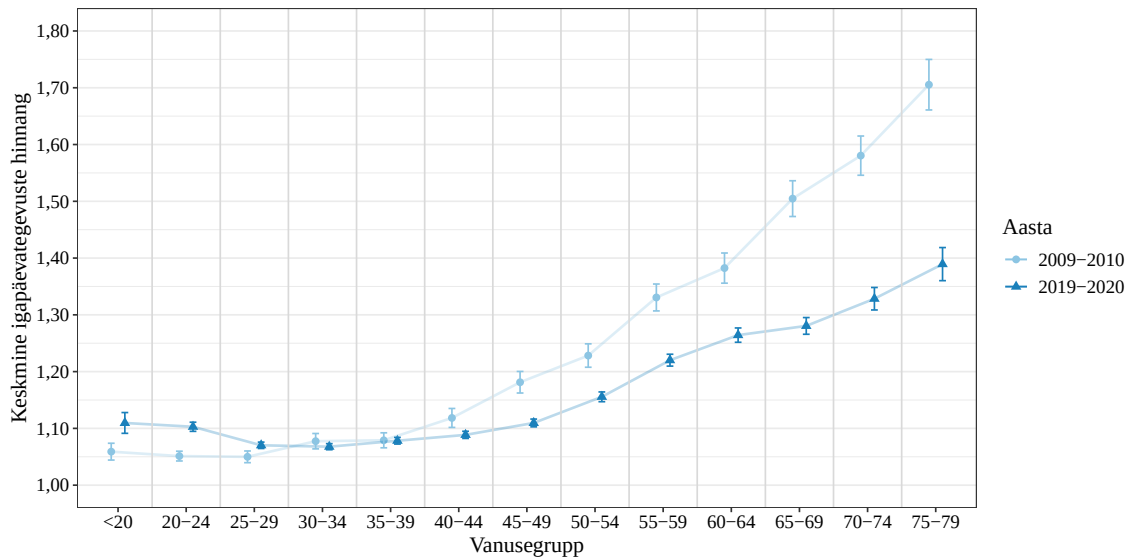
Joonis 4: Keskmise liikuvuse dimensiooni hinnang vanusegruppides koos 95%-usaldusvahemikega

Enesehoolitsuse keskmised hinnangud olid vahemikus 1 – 1,27. Keskmiste erinevus oli oluline alates 50–54-aastaste vanusegrupist ning sarnaselt liikuvuse hinnangule, oli ka seal edaspidi keskmine madalam aastatel 2019–2020 ning erinevus suurenes vanuse kasvades. Nooremate vanusegruppide hinnangute sarnasus on samuti eeldatav, sest hinnangud olid varemgi väga ühelähedased. (joonis 5)



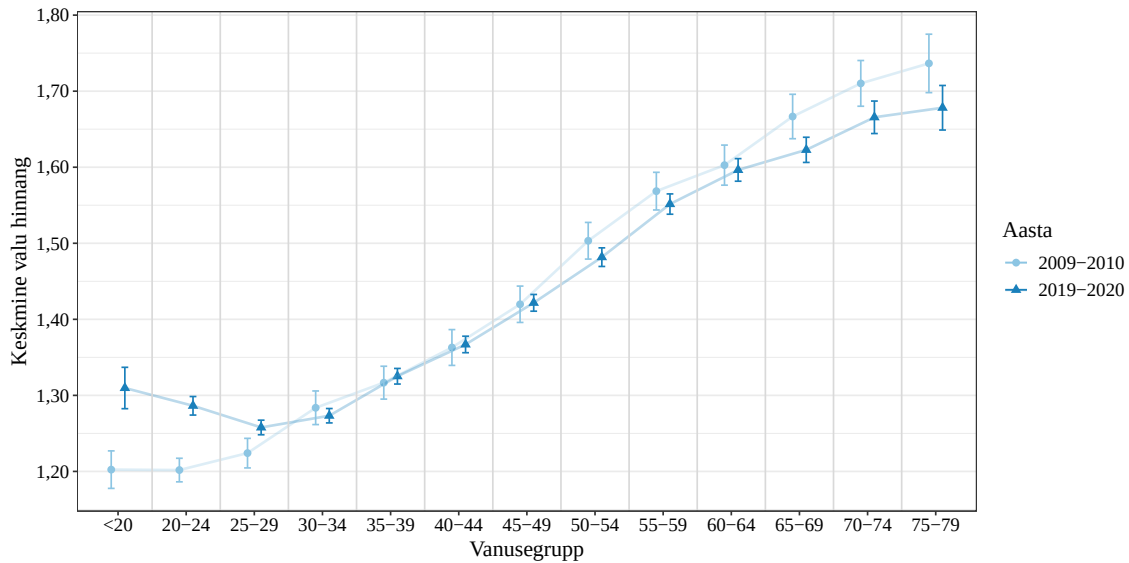
Joonis 5: Keskmise enesehoolitsuse dimensiooni hinnang vanusegruppides koos 95%-usaldusvahemikega

Igapäevategevustega toimetulekut hindasid aastatel 2019–2020 nooremad vanusegrupid (kuni 24-aastased) halvemini kui 10 aastat varem, kuid erinevus ei olnud väga suur – kuni 0,09 ühikut. Vanemate vanusegruppide (alates 40-aastaste) seas olid aga probleemid vähenenud ning muutuse määr suurenes taas koos vanuse kasvuga. Hinnangud jäid üldjoontes vahemikku 1,04–1,75 ehk erinesid kuni 0,71 ühikut. (joonis 6)



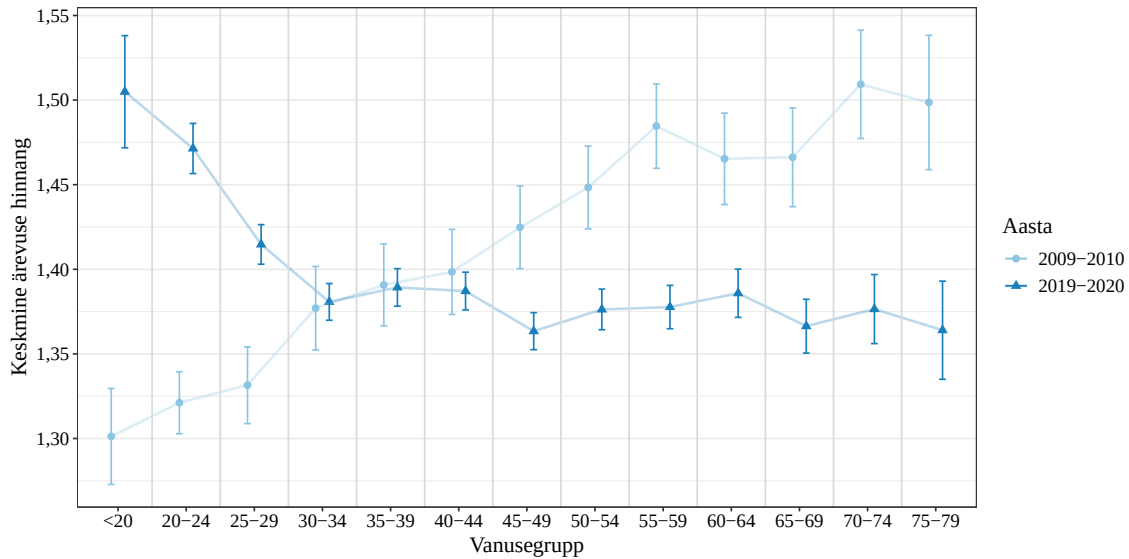
Joonis 6: Keskmine igapäevategevustega toimetuleku dimensiooni hinnang vanusegruppides koos 95%-usaldusvahemikega

Valu hinnangud olid vahemikus 1,18–1,77. Võrreldes eelnevate dimensioonide keskmiste erinevustega kohortide vahel, on valu hinnangutes enim vanusegruppe, kus hinnang ei olnud kümne aastaga oluliselt muutunud. 2019.–2020. aastate kohordis oli hinnang kõrgem kuni 24-aastateni ning ülejäänud vanusegruppides ei olnud toimunud olulisi muutusi. (joonis 7)



Joonis 7: Keskmine valu dimensiooni hinnang vanusegruppides koos 95%-usaldusvahemikega

Ärevuse ja depressiooni dimensioonis olid keskmiste erinevused vanusegruppides kohortide vahel kõige suuremad. Kuni 25–29-aastaste vanusegrupini hindasid oma ärevust kõrgemalt 2019.–2020. aastate geenidonorid ning erinevus kümme aastat varasemate hinnangutega jäi vahemikku 0,14 – 0,27. Kui aastatel 2009–2010 kasvas ärevuse hinnang kuni 29-aastasteni, siis hilisemas kohordis muutus see vastupidi – selles vahemikus vanuse suurenedes vähenesid probleemid ärevusega. 30–44-aastastel ei olnud toimunud keskmistes hinnangutes olulisi muutusi, kuid alates 45-aastastest oli sarnane muster nagu kuni 29-aastastel ehk mõlema kohordi hinnangud muutuvad vastassuundades, kuid siin oli parem hinnang 2019.–2020. aastate geenidonoritel. (joonis 8)



Joonis 8: Keskmine ärevuse dimensiooni hinnang vanusegruppides koos 95%-usaldusvahemikega

Võrreldes varasema kohordiga, on aastatel 2019–2020 näha keskmistes dimensiooni hinnangutes nii tõuse kui ka langusi. Vanemate doonorite hinnangud olid üldjoontes paranenud või jäänud samaks igas dimensioonis, kusjuures vanusegrupi kasvades suurenes ka muutuse määr. Kuni 24-aastastel oli halvenenud hinnangud nii igapäevatoimetuste, valu kui ka ärevuse dimensioonides, seejuures ärevuse dimensiooni hinnangute erinevus kahe kohordi noorimate vahel oli kõigi dimensioonide suurim. Kõige vähem muutusi kahe kohordi vanusegruppide keskmiste hinnangute vahel ilmes valu dimensioonis.

3.2.2 K-keskmiste klasteranalüüs

Kasutades klasteritunnustena dimensioonide hinnanguid ning rakendades optimaalse klastrite arvu valimiseks küünarnuki meetodit (lisa 2), saab moodustada neli klastrit. Tabelis 6 on esitatud saadud klastrite keskmised, mille põhjal saame neid iseloomustada järgmiste gruppidega (järjestatud kõige tervemast kõige „hullemani“):

1. vähesed probleemid kõigis dimensioonides (keskmised hinnangud ~ 1),
2. mõningad probleemid valu ja ärevusega (keskmised hinnangud ligikaudu vahemikus $\sim 1,5\text{--}2$),

3. mõningad probleemid liikuvuse, igapäevaste toimetuste, valu ja ärevusega,
4. mõningad probleemid ärevusega ja suured probleemid kõigis teistes dimensioonides (suured probleemid, kui keskmine hinnang on suurem kui 2).

Tabel 6: 2009.–2010. aastate kohordi klasterikeskmised

Dimensioon/ Klaster	1	2	3	4
Liikuvus	1,01	1,00	1,99	2,02
Enesehoolitsus	1,00	1,00	1,01	2,07
Igapäevatoimetused	1,01	1,35	1,77	2,30
Valu	1,12	1,97	1,93	2,06
Ärevus	1,18	1,96	1,65	1,81

Arvestades, et klasterid on järjestatud parimast kuni halvima üldise tervisehinnangu, mida kinnitab ka klasterite keskmine LSS väärtus, on tabelist 7 näha, et see käib käsikäes ka vanusega – keskmine vanus oli madalaim esimeses klasteris (39,75 aastat) ja kõrgeim neljandas (66,98 aastat). Sooline jaotus oli klasterites üldjoontes sarnane kogu valimi jaotusele – esimeses kohordis oli 62,80% naisi ning ka kõigis klasterites jäi see osakaal vahemikku 61–67%. Haridusastmete osakaalude muutuses esines aga eristatav muster – keskmise tervisehinnangu halvenedes suurenes põhihariduse ja madalama haridusega doonorite ning vähenes kõrgharidusega isikute osakaal.

Tabel 7: 2009.–2010. aastate kohordi klasterite ülevaade

Tunnus	Klaster 1	Klaster 2	Klaster 3	Klaster 4
Suurus	13 831	3864	2486	1338
Keskmine vanus	39,75	46,88	60,80	66,98
Vanuse standardhälve	16,20	16,20	15,1	14,50
Mehi (%)	38,97	33,00	34,59	35,95
Naisi (%)	61,03	67,00	65,41	64,05
Haridustase (%)				
Algharidus	6,33	10,66	22,33	35,65
Põhiharidus	31,91	39,52	38,50	37,00
Kõrgharidus	61,76	49,82	39,18	27,35
Keskmine LSS	5,32	7,28	8,35	10,26

Ka teise kohordi puhul otsustati moodustada neli klasterit (lisa 2), mida saab iseloomustada tabelis 8 esitatud keskmiste põhjal järgmiste gruppidega (järjestatud kõige tervemast kõige „hullemani“):

1. vähesed probleemid kõigis dimensioonides,
2. suured probleemid ärevusega,
3. mõningad probleemid ärevusega ja suured probleemid valuga,
4. mõningad probleemid liikuvuse ja ärevusega ning suured probleemid igapäevatoimetuste ja valuga.

Tabel 8: 2019.–2020. aastate kohordi klasterikeskmised

Dimensioon/ Klaster	1	2	3	4
Liikuvus	1,01	1,01	1,12	1,65
Enesehoolitsus	1,00	1,00	1,00	1,21
Igapäevatoimetused	1,01	1,06	1,00	2,03
Valu	1,00	1,03	2,01	2,07
Ärevus	1,00	2,07	1,47	1,72

Teise kohordi puhul oli esimese klasteri keskmine vanus kõrgem kui teises klasteris – noorima keskmise vanusega oli klaster, mille põhiliseks terviseprobleemiks oli ärevus ja depressioon. Meeste ja naiste osakaal igas klasteris jäi taaskordselt sarnaseks valimi üldisele jaotusele ning haridustasemete osakaalude muutumise muster sarnanes eelmises kohordis täheldatule. Võrreldes 2009.–2010. aastate kohordiga, oli keskmine LSS väärtus igas klasteris mingil määral vähenenud, aga enim just kõrgema keskmise vanusega klasterites – suurim muutus oli kolmandas klasteris, kus, võrreldes esimese kohordiga, oli vähenenud keskmine LSS 1,74 võrra ehk robustselt hinnates oli paranenud keskmine profiil 1–2 probleemitaseme võrra. (tabel 9)

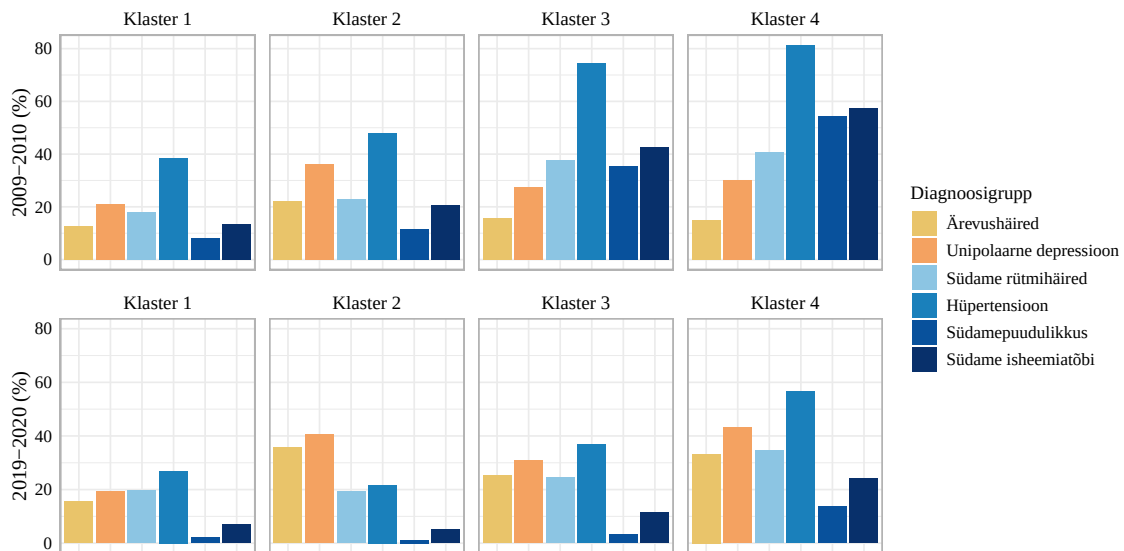
Tabel 9: 2019.–2020. aastate kohordi klasterite ülevaade

Tunnus	Klaster 1	Klaster 2	Klaster 3	Klaster 4
Suurus	33897	12585	20779	9467
Keskmine vanus	41,69	38,21	46,32	53,73
Vanuse standardhälve	13,82	13,15	14,67	15,33
Mehi (%)	33,95	29,49	29,49	29,52
Naissi (%)	66,05	70,51	70,51	70,48
Haridustase (%)				
Põhiharidus või madalam	1,94	2,41	3,10	6,44
Keskkharidus	25,56	27,06	30,62	38,88
Kõrgharidus	72,49	70,53	66,28	54,67
Keskmine LSS	5,02	6,17	6,61	8,68

2003. aasta Eesti rahvastiku tervisekaotuse uuringust leiti, millistel diagnoosigrupidel oli suurim mõju tervisekaotusele. Nende tulemuste põhjal valiti välja kuus diagnoosigruppi, mille esinemisi klastrites lähemalt uuriti. Diagnoosigrupid valiti selle järgi, millised neist kuulusid kõige sagedamini esinenud diagnooside hulka võimalikult paljudes klastrites. Diagnoosigrupid on järgmised:

- südame isheemiatõbi,
- südame rütmihäired,
- südamepuudulikkus,
- hüpertensioon,
- ärevushäired,
- unipolaarne depressioon.

Gruppidesse kuuluvate täpsemate diagnooside loetelu on esitatud lisas 4.



Joonis 9: Diagnoosigruppide jaotused klastrites

Võttes vaatluse alla esimese kohordi diagnoosigruppide jaotuse ning klastrit iseloomustavad tervises seisundid, näeme, et klastrites, mille korral oli vähemalt mõningaid probleeme liikuvuse, igapäevatoimetuste, valu ning ärevusega (klastrid 3 ja 4), esines

rohkem ka südamehaigustega seotud diagnoose. Arvestades, et kolmas klaster sisaldas peamiselt indiviide, kes täheldasid mõningaid probleeme kõigis dimensioonides peale enesehoolitsuse, ning neljas klaster neid, kes täheldasid suuri probleeme samades dimensioonides, ilmneb, et suurim erinevus nende klastrite diagnooside jaotuses on seotud isheemiatõve ja südamepuudulikkuse sageduse suurenemisega. Klastrites, kus täheldati probleeme ärevusega (klastrid 2, 3 ja 4) on ootuspäraselt ka suurem ärevuse ja depressioonimäär, seejuures olid osakaalud suurimad just teises klastris, kus ka hinnati seda keskmiselt kõige kõrgemalt (1,96 vs 1,65 ja 1,81).

Teises kohordis olid kõigis klastrites diagnooside osakaalud märgatavalt väiksemad. Olulisel määral oli vähenenud südamepuudulikkus, mille esinemine just esimeses, teises ja kolmandas klastris oli väga väike. Sarnaselt eelneva kohordiga ilmnis seos liikuvuse ja igapäevatoimetuste hinnangute ning südamehaiguste osakaalu vahel – võrreldes kolmanda klastriga oli neljandas klastris rohkem probleeme liikuvuse ning igapäevatoimetuste dimensioonides ning diagnooside jaotuses olid neljandas klastris ka sagedamad isheemiatõbi ning südamepuudulikkus. Seejuures oli neljandas klastris lisaks nendele märgatavalt kõrgem ka depressiooni ning ärevushäirete diagnooside osakaal. Arvestades, et teises klastris oli ärevuse hinnangu keskmine 2,07 ning neljandas 1,72, on märkimisväärne, et mõlemas klastris olid depressioonidiagnooside ning ärevushäirete osakaalud ligikaudu sama suured, mis võib viidata sellele, et individid, kellel esineb teisi somaatilisi terviseprobleeme, ei hinda oma ärevust nii äärmuslikult, kui seda teevad inimesed, kellel neid ei ole või kellel neid on vähem.

Uurides mõlema kohordi klastritesisest vanuselist jaotust, mis on leitav lisast 3, selgub, et esimese kohordi kolmanda ja neljanda klatri moodustasid suuremas osas vastavalt üle 50 ning üle 80-aastased doonorid. Selline vanuseline jaotus aga ei tekkinud mitte üheski 2019.–2020. aastate kohordi klastris, mis võib olla tingitud nii kohortide vanuselise jaotuse erinevustest – esimeses kohordis oli märgatavalt rohkem üle 70-aastaseid – kui ka üldiste hinnangute paranemisest kümme aastat hilisemas kohordis just vanemas eas, mida täheldasime ka eelnevas alapeatükis. Seetõttu lähtuti kohortide klastritevahelisel võrdlusel vanuse jaotuse sarnasusest, kuna klastrite numeratsioon ei peegelda hästi võrreldavaid vanusestruktuure. Võrreldavateks klastriteks osutusid

- (1) 2009–2010 esimene klaster ja 2019–2020 teine klaster, kuna mõlema puhul oli suurim osakaal koondunud vanusegrupi 25–29 ümber.
- (2) 2009–2010 teine klaster ja 2019–2020 kolmas klaster, kuna suurim osakaal koondus vanusegruppide 45–49 ja 50–54 ümber.

Arvestades, et võrdlusesse (1) kuuluvad mõlema kohordi noorimad klastrid (keskmised vanused vastavalt 39,75 ja 38,21), siis on märgata, et nooremate seas olid sagedamini ärevuse ja depressiooni diagnoosid aastatel 2019–2020, kuid märkimisväärselt vähenenud hüpertensioon, isheemiatõbi kui ka südamepuudlikkus. Oluline on välja tuua, et vaatamata sellele, et võrreldakse sarnase vanuselise profiiliga klastreid, ei saa välistada, et erinevus ei ole seotud vanuselise jaotuse erinevustega – aastatel 2009–2010 on osakaalud vanusegruppides ühtlasemad, aga hilisemas kohordis domineerivad rohkem 25–44-aastased, mistõttu on võimalik, et ka varasemas kohordis esines samal määral ärevushäireid nendes vanusegruppides, kuid nende kõrgeid hinnanguid tasakaalustasid teiste vanusegruppide suuremad osakaalud.

Võrdlusesse (2) kuuluvates klasterites on vähesel määral vähenenud depressiooni diagnoosid ning tõusnud ärevushäirete diagnoosid aastatel 2019–2020, kuigi varasemas kohordis hindasid sellesse klasterisse kuuluvad indiviidid oma ärevust märgatavalt kõrgemalt (1,96 vs 1,47). Vähenenud olid aga nii isheemiatõve kui ka südamepuudlikkuse diagnooside esinemised.

3.3 EuroQoli indeksid

Käesolevas peatükis uuritakse kohortide EuroQoli indekseid, mis arvutati W. Greineri poolt 2003. aastal koostatud Euroopa *value set*'i alusel (lisa 6). Indeks esitab tervisega seotud elukvaliteeti koondnäitajana ning selle suurem väärtus on seotud kõrgema tervisega seotud elukvaliteediga.

Aastatel 2009–2010 oli keskmine indeksi väärtus 0,815, mis on 0,018 võrra väiksem 2019.–2020. aastate keskmisest indeksist (0,833). Märkimisväärne on asjaolu, et naised hindavad oma elukvaliteeti aastatel 2019–2020 samal tasemel nagu tegid seda mehed 10 aastat varem, seejuures meeste tervisega seotud elukvaliteet on mõlemas

kohordis naiste omast kõrgem. Mediaan aga on nii kohortide kui ka sugude lõikes sama – 0,783. (tabel 10)

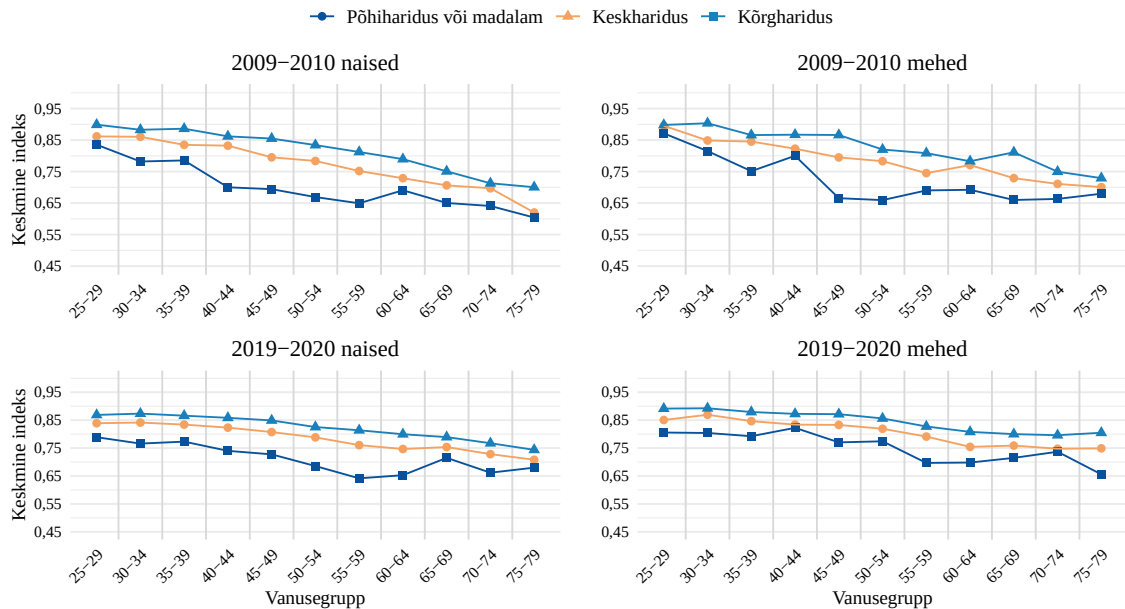
Tabel 10: EuroQoli indeksi jaotust kirjeldavad näitajad

Tunnus	2009–2010	2019–2020
Keskmine (95%-UI)	0,815 (0,812, 0,818)	0,833 (0,832, 0,834)
Naised	0,808 (0,805, 0,811)	0,827 (0,826, 0,829)
Mehed	0,827 (0,823, 0,831)	0,845 (0,843, 0,848)
Mediaan	0,783	0,783

Koostades neli dispersioonanalüüsi mudelit kohordi ja soo lõikes, kus argumenttunnusteks on vanusegrupp ning haridusaste, selgub, et vanusegrupi ja haridusastme koosmõju on kõigis mudelites, välja arvatud 2019–2020 meeste korral, oluline. Mudelitest jäeti välja alla 25-aastased, sest nende seas on kõrgahariduse omandanud väga vähe, ja üle 80-aastased varasemalt mainitud kohortide osakaalude erinevuse tõttu. Tulemused on leitavad lisast 7.

Lähtuvalt mudelites täheldatud koosmõju olulisusest, uuriti täpsemalt keskmise indeksi väärtuse muutust vanusegruppides haridusastmete lõikes. Ootuspäraselt väheneb indeksi väärtus kõigi haridusastmete korral vanuse kasvades. Kõrgharidusega geenidonoritel on mõlemas kohordis nii meeste kui ka naiste seas kõrgeim keskmine indeks, sellele järgneb keskharidus ning kõige madalam keskmine indeks on üldjoontes põhihariduse või madalama haridusega geenidonoritel. Jooniselt 10 on näha, et mõnevõrra erandlikult muutub põhihariduse või sellest madalama haridusastmega indiviidide keskmine indeks vanusegruppides võrreldes kõrgemate haridusastmetega. Kõrg- ning keskharidusastmete korral on enamjaolt selge suundumus – mida vanem on indiid, seda madalam on indeksi väärtus ning see langeb ühtlaselt vanuse suurenedes. Põhihariduse korral aga tõuseb ja langeb keskmine indeksi väärtus rohkem vanusegruppide vahel. Selle haridusega meeste seas joonistub välja ajas nihkuv muster: perioodil 2009–2010 toimub indeksi märgatav langus vanuserühmades 35–39, 45–49 ja 65–69. Kümme aastat hiljem, perioodil 2019–2020, esineb langus vastavalt kümme aastat vanemates vanuserühmades, st 45–49, 55–59 ja 75–79-aastastel, kuid langused on väiksemad, mis võib viidata mingit laadi generatsioonilisele eripärale. Naistel sellist mustrit aga ei teki, kuid 2019.–2020. aastatel on põhiharidusega

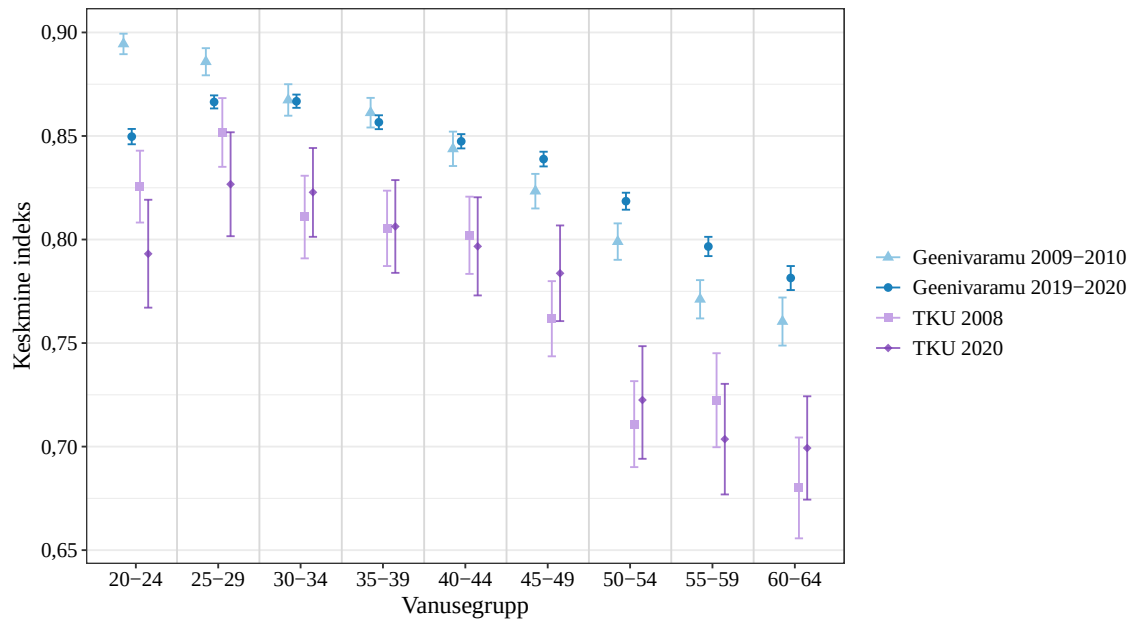
naiste keskmise indeksi langus ühtlasem kui varem ning langeb kuni 55–59-aastaste vanusegrupini.



Joonis 10: Keskmise indeksi väärtus vanusegruppide ja haridusastmete lõikes

3.3.1 Võrdlus täiskasvanud elanikkonnaga

Võttes vaatluse alla Eesti täiskasvanud rahvastiku tervisekäitumise uuringu (TKU) aastatel 2008 ja 2020, on näha, et TKU keskmised EuroQoli indeksi väärtused ei erine teineteisest oluliselt üheski vanusegrupis, sest usaldusintervallid kattuvad. Üldiselt on geenidoonorite hinnang oma elukvaliteedile parem kui täiskasvanud elanikkonna juhuvalimis. Geenidoonorite keskmisi indeksi väärtusi omavahel võrreldes on aga näha, et keskmised ei erine (usaldusvahemike põhjal) vaid vanusegruppides 30–34, 35–39 ja 40–44. Samuti esineb varem dimensioonide keskmisi hinnanguid uurides ilmsiks tulnud muster seoses muutustega kohortide vanusegruppides – 2019.–2020. aastatel on nooremad vanusegrupid (20–24 ja 25–29) hinnanud oma elukvaliteeti halvemini, kuid vanemates vanusegruppides (45-aastased ja vanemad) on hinnangud paranenud. (joonis 11)



Joonis 11: TKU ja geenivaramu kohortide keskmised indeksi väärtused vanusegruppides

Arutelu

Kahe geenidoonorite kohordi võrdluses näeme üldiselt kõrgemat elukvaliteeti hili-semas ehk 2019.–2020. aastate kohordis, aga see erinevus ei ole igas EuroQoli kü-simustiku poolt mõõdetavas dimensioonis ühesugune. Nii aastatel 2009–2010 kui ka 2019–2020 olid sagedasemad profiilid, mille korral täheldati mõningaid problee-me valu ja/või ärevusega, kuid elukvaliteedi tõus väljendus profiilide põhjal enim muutustes seoses liikuvuse, enesehoolitsuse ja igapäevatoimetustega. Shannoni in-deksite põhjal selgub, et hinnangute mitmekülgsus oli nendes kolmes dimensioonis vähenenud ning osakaalude muutus kahe kohordi hinnangute jaotuste vahel kinni-tab, et vähenenud olid nii mõningate kui ka väga suurte probleemide hinnangud, mille arvelt olid aga tõusnud väheste probleemide hinnangud. Valu dimensioonis oli muutus väike ning väheste probleemidega indiviidide osakaal oli aastatel 2019–2020 vaid mõnevõrra suurenenud. Seevastu aga ärevuse dimensioonis oli küll vähenenud mõningate probleemidega doonorite osakaal, kuid selle arvelt tõusnud nii väheste kui ka väga suurte probleemide hinnangute osakaal.

Ka klasteranalüüs viitab sellele, et liikuvuse ja igapäevategevustega toimetulekuga seotud probleemid olid teises kohordis vähem esinevad. Kui esimeses kohordis moo-dustus klaster, mis hõlmas samaaegselt probleeme liikuvuse, igapäevatoimetuste, valu ja ärevusega, siis teises kohordis sellist klastrit enam ei tekkinud. Selle asemel moodustus 2019.–2020. aastate kohordis aga klaster, mis iseloomustas indiviide, kes täheldasid väga suuri probleeme vaid ärevusega, seejuures oli selle keskmine vanus noorem kui klastris, kus hinnati kõigis dimensioonides väheseid probleeme. Klast-rites, kus esines probleeme ärevusega, esines ka rohkem ärevushäirete ning depres-siooni diagnoose ning aastatel 2019–2020 olid nende diagnooside osakaalud suure-mad. Seejuures viitas diagnooside jaotus võimalikule seosele, et indiviidid, kellel esines rohkem somaatilisi terviseprobleeme, ei hinnanud oma ärevuse taset nii kõr-gelt nagu seda tegid need, kellel selliseid probleeme nii palju ei olnud. Arvestades, et 2019.–2020. aastatel olid suurel määral vähenenud vaadeldud südamehaigused, võib see olla eelneva tähelepaneku põhjal seotud ärevusprobleemide parema tajumisega. Seejuures on vaadeldava perioodi ajal hakatud ka ühiskonnas rohkem vaimse tervise

probleemidele tähelepanu pöörama, millega võib kaasneda nende probleemide parem teadvustamine.

Lisaks selgub, et hinnangute erinevuse suund kui ka ulatus kahe kohordi vahel sõltus tugevalt vanusest. Nii liikuvuse, enesehoolitsuse kui ka igapäevategevustega toimetuleku keskmised hinnangud olid aastatel 2009–2010 nooremate seas ühelähedased ehk täheldati väheseid probleeme ning kümme aastat hiljem need oluliselt muutunud ei olnud. Kuni 24-aastaste seas olid suurenenud probleemid hilisemas kohordis valu ja ärevuse dimensioonides, seejuures ärevuse hinnangud olid halvenenud kõige enam. Samas olid peaaegu igas dimensioonis keskmised hinnangud paranenud 45-aastastest alates, kus ainsaks erandiks oli valu, mille hinnangutes ei olnud üheski vanemas vanusegrupis olulisi muutusi toimunud.

Rakendades profiilidele EuroQoli Euroopa *value set*'i selgub, et 2009.–2010. aastate meeste keskmine indeksi väärtus oli võrdne 2019.–2020. aastate naiste keskmise indeksiga ehk hilisema kohordi naiste tervisega seotud elukvaliteet oli samal tasemel nagu meestel kümme aastat varem. Seejuures olid mõlemas kohordis meeste keskmised indeksid kõrgemad naiste omadest. Kõrg- ja keskharidusega indiviidide keskmise indeksi muutus vanuse kasvades nii naiste kui ka meeste seas oli sarnane, seejuures keskmine oli alati kõrgeim kõrgharitudel. Seda asjaolu, et kõrgema haridusega indiviidid hindasid oma elukvaliteeti kõrgemalt, täheldati ka klasteranalüüsis, kus parema tervisehinnanguga klastrites suurenes kõrgharitude osakaal. Kõrgem haridus võib tähendada paremat terviseteadlikkust kui ka viidata paremale ravi kättesaadavusele – OECD uuringus, kuhu kuulus ka Eesti, täheldati, et kõrgem haridusaste on seotud nii väiksema töötuse kui ka kõrgema palgaga (Haridus- ja Teadusministeerium, 2025). Põhiharidusega geenidoonorite keskmine indeks oli aga madalaim, seejuures ei muutunud see vanuse kasvades nii ühtlaselt nagu kõrgema haridusega indiviididel.

Geenidoonorite tervisega seotud elukvaliteet oli 2009.–2010. aastate kohordis kõrgem 2008. aastal Eesti täiskasvanud rahvastiku tervisekäitumise uuringus täheldatud tulemustest ning aastatel 2019–2020 kõrgem sama uuringu 2020. aasta tulemustest. Arvestades, et ka TKU tulemustes võis ilmnedasarnane muster indeksiväärtus-

te muutuses nagu geenidoonoritel, kuid antud tulemuste korral ei ole see laiadest usaldusintervallides tulenevalt tõestatav, võivad peegeldada geenidoonorite tulemusel mingil määral elukvaliteedi muutust kogu elanikkonnas. Tuleb aga arvestada asjaoluga, et geenidoonorid ei ole juhuvalim ning kaasamise meetoditest tulenevalt on võimalik, et aastatel 2019–2020 liitunud geenidoonorid esindavad tervemaid inimesi, võrreldes aastatel 2009–2010 liitunudega. Arvestades, et eelnevalt täheldati 2019.–2020. aastate kohordi noorematel mõndades dimensioonides halvemaid hinnanguid ning vanematel enamikes dimensioonides paremaid tulemusi, ei ole üllatav, et see muster esineb ka indeksi muutuses.

Kokkuvõte

Bakalaureusetöö eesmärgiks oli välja selgitada, milline on Tartu Ülikooli Eesti geenivaramu doonorite tervisega seotud elukvaliteet, kuidas see on muutunud kümne aasta jooksul ning millised seosed on sellel erinevate demograafiliste ja kliiniliste teguritega. Analüüs viidi läbi aastatel 2009–2010 ja 2019–2020 liitunud geenidoonorite andmetel. Kasutatavad meetodid sisaldasid erinevaid EuroQoli küsimustiku tulemuste jaotusi iseloomustavaid näitajaid, K-keskmiste klasteranalüüsi ning keskmiste mitmest võrdlemist.

2019.–2020. aastate kohordis oli geenidoonorite tervisega seotud elukvaliteet kõrgem nii meestel kui ka naistel võrreldes 2009.–2010. aastate kohordiga, kuid mõlema kohordi korral oli keskmine EuroQoli indeks kõrgem just meestel. Põhiliselt olid vähenenud probleemid liikuvuse, enesehoolitsuse ning igapäevategevustega toimetuleku dimensioonides ja seda eriti vanemate doonorite seas. Probleemid valuga olid jäänud aastatel 2019–2020 samale tasemele nagu need olid kümme aastat varasemas kohordis. Noorematel oli hilisemas kohordis tekkinud rohkem probleeme ärevusega, mis olid sama kohordi eakamatel aga vähenenud 2009.–2010. aastate kohordiga võrreldes. Ärevuse probleemide kasvule viitab ka ärevuse ning depressiooni diagnooside sagenemine hilisemas kohordis, seejuures inividid, kellel oli vähemalt mõningad probleeme ärevuse ja depressiooniga, olid suurema tõenäosusega vastava diagnoosi kunagi ka saanud. Lisaks selgus, et haridusel oli mõju EuroQoli indeksi väärtusele – põhihariduse või madalama haridusega inividide indeksi väärtuse muutus vanusegrupiti oli erinev võrreldes kesk- ja kõrgharidusega ning ei käitunud nii ühtlaselt vanusegruppide lõikes. Seejuures oli kõrgeim keskmine indeks igas vanusegrupis kõrgharitudel ning madalaim põhihariduse või madalama haridusega inivididel. Võrreldes aga geenidoonorite indeksite väärtusi Eesti täiskasvanud elanikkonna tervisekäitumise 2008. ja 2020. aasta tulemustega, selgus, et geenidoonorite tervisega seotud elukvaliteet oli üldiselt kõrgem täiskasvanud elanikkonna juhuvalimist kõigis vanusegruppides.

Töö edasiarenduseks oleks otstarbekas teha järjestustunnuste mudeleid dimensioonidele nagu näiteks võrdeliste šansside mudel, mille abil saaks täpsemalt hinnata

erinevate tegurite mõju dimensiooni hinnangule. Seejuures oleks võimalus võtta arvesse ka geenivaramu kohortide võimalikku valikunihet, mis on tingitud erinevast doonorite kaasamise meetodist. Detailsemalt saaks võrrelda geenidoonorite tervisega seotud elukvaliteeti ka täiskasvanud rahvastiku tervisekäitumise uuringu tulemustega, analüüsides erinevusi dimensioonide hinnangute tasemel. Samuti võiks huvi pakkuda, kas ja kui suurel määral aitab elukvaliteedi hinnang prognoosida tulevikus tekkivaid terviseriske või oodatavat eluiga.

Kasutatud allikad

- Bland, J. M. ja D. G. Altman (1995). "Multiple significance tests: the Bonferroni method". *BMJ* 310.6973, lk. 170. ISSN: 0959-8138. DOI: [10.1136/bmj.310.6973.170](https://doi.org/10.1136/bmj.310.6973.170). eprint: <https://www.bmj.com/content/310/6973/170.full.pdf>. URL: <https://www.bmj.com/content/310/6973/170>.
- Devlin, N., D. Parkin ja B. Janssen (2020). *Methods for analysing and reporting EQ-5D data*. Springer. DOI: [978-3-030-47622-9](https://doi.org/10.1007/978-3-030-47622-9).
- Haqiqatkhah, M. M. (2025). "Floor and ceiling effects". *MATILDA*.
- Haridus- ja Teadusministeerium (9. september 2025). *OECD haridusstatistika: mida kõrgem haridus, seda suurem palk ja väiksem töötuse risk*. URL: <https://www.hm.ee/uudised/oecd-haridusstatistika-mida-korgem-haridus-seda-suurem-palk-ja-vaiksem-tootuse-risk> (vaadatud 09.05.2026).
- Hartigan, J. A. ja M. A. Wong (1979). "Algorithm AS 136: K-Means clustering algorithm". *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, lk. 100–108. DOI: [10.2307/2346830](https://doi.org/10.2307/2346830).
- James, G., D. Witten, T. Hastie ja R. Tibshirani (2023). *An introduction to statistical learning with applications in R*. Springer. DOI: [978-3-030-47622-9](https://doi.org/10.1007/978-3-030-47622-9).
- Janssen, M. F., E. Birnie ja G. J. Bonsel (2007). "Evaluating the discriminatory power of EQ-5D, HUI2 and HUI3 in a US general population survey using Shannon's indices". *Qual Life Res* 16, lk. 895–904. DOI: [10.1007/s11136-006-9160-6](https://doi.org/10.1007/s11136-006-9160-6).
- Karimi, M. ja J. Brazier (2016). "Health, health-related quality of life, and quality of life: what is the difference?" *Pharmacoeconomics* 34.7, lk. 645–649. DOI: [10.1007/s40273-016-0389-9](https://doi.org/10.1007/s40273-016-0389-9).
- Kodinariya, T. ja P. Makwana (2013). "Review on Determining of Cluster in K-means Clustering". *International Journal of Advance Research in Computer Science and Management Studies* 1, lk. 90–95.
- Käärik, E. (s.a.). *Andmeanalüüs II*. Loengukonspekt, Tartu Ülikool.
- Lai, T. (2012). *Ülevaade peamistest üldistest terviseiga seotud elukvaliteeti mõõtvatest küsimustikest ja nende kasutamisest Eestis*. URL: <http://hdl.handle.net/10062/85318>.

- Lember, J. (2022). *Informatsiooniteooria*. Loengukonspekt ja ülesanded, Tartu Ülikool.
- Matter-Walstra, K., D. Klingbiel, T. Szucs, B. C. Pestalozzi ja M. Schwenkglenks (2014). “Using the EuroQol EQ-5D in Swiss Cancer patients, which value set should be applies?” *Pharmacoeconomics* 32, lk. 591–599. DOI: [s40273-014-0151-0](https://doi.org/10.1007/s40273-014-0151-0).
- Sarty, Gordon E. (2022). *Introduction to Applied Statistics for Psychology Students*. University of Saskatchewan Open Press. URL: <https://www.saskoer.ca/introtoappliedstatsforpsych/>.

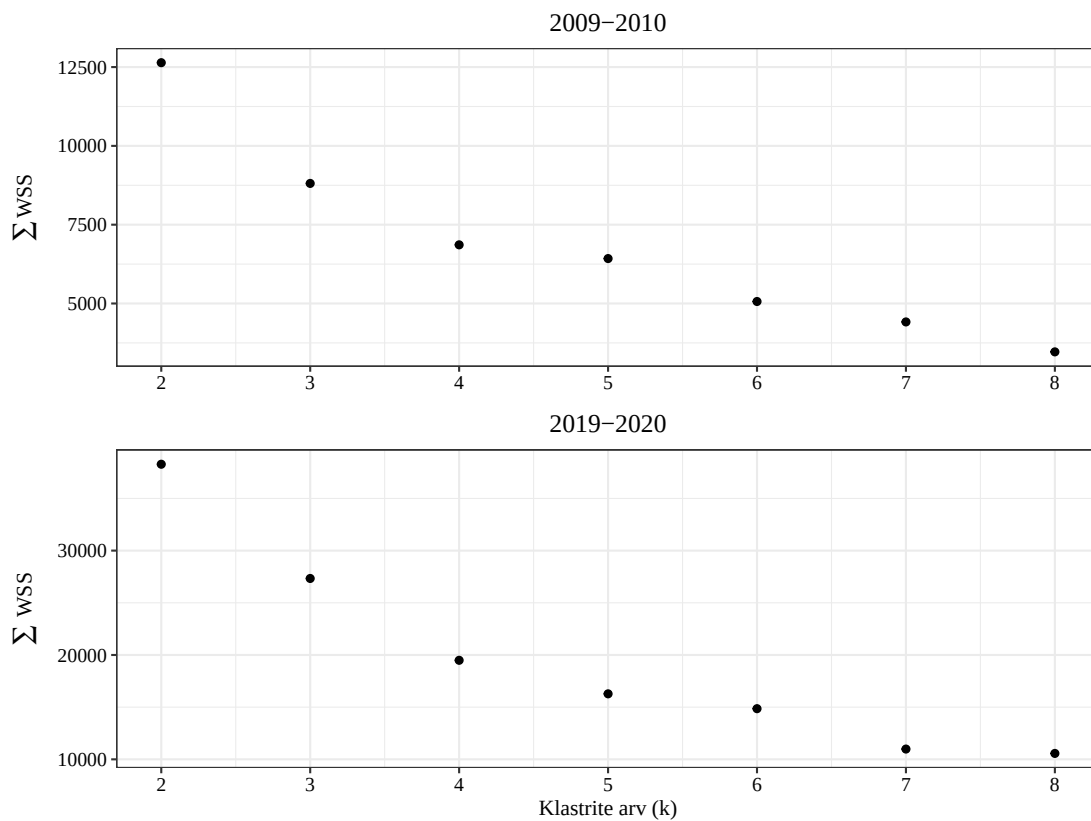
Lisad

Lisa 1: Dimensioonide hinnangute tasemete osakaalud kohortides

Tabel L1: Dimensioonide hinnangute tasemete jaotus kohortide kaupa. Iga dimensiooni puhul summeeruvad osakaalud vastavas veerus 100%-ks.

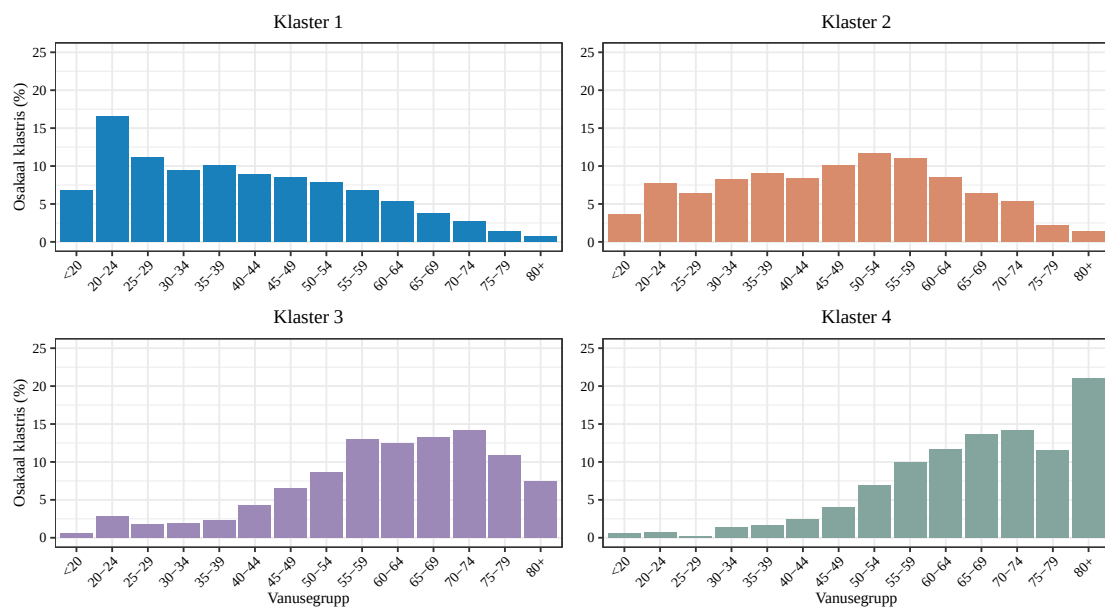
Dimensioon		2009–2010	2019–2020
Liikuvus (%):	1	82,09	88,02
	2	17,46	11,93
	3	0,45	0,05
Enesehoolitsus (%):	1	93,68	97,37
	2	5,80	2,54
	3	0,52	0,09
Igapäevaste tegevustega toimetulek (%):	1	78,33	86,34
	2	19,41	13,23
	3	2,26	0,43
Valu (%):	1	58,78	60,40
	2	39,85	38,23
	3	1,37	1,37
Ärevus/depressioon (%):	1	60,47	62,94
	2	37,80	34,96
	3	1,73	2,09

Lisa 2: Küünarnuki meetodi rakendus

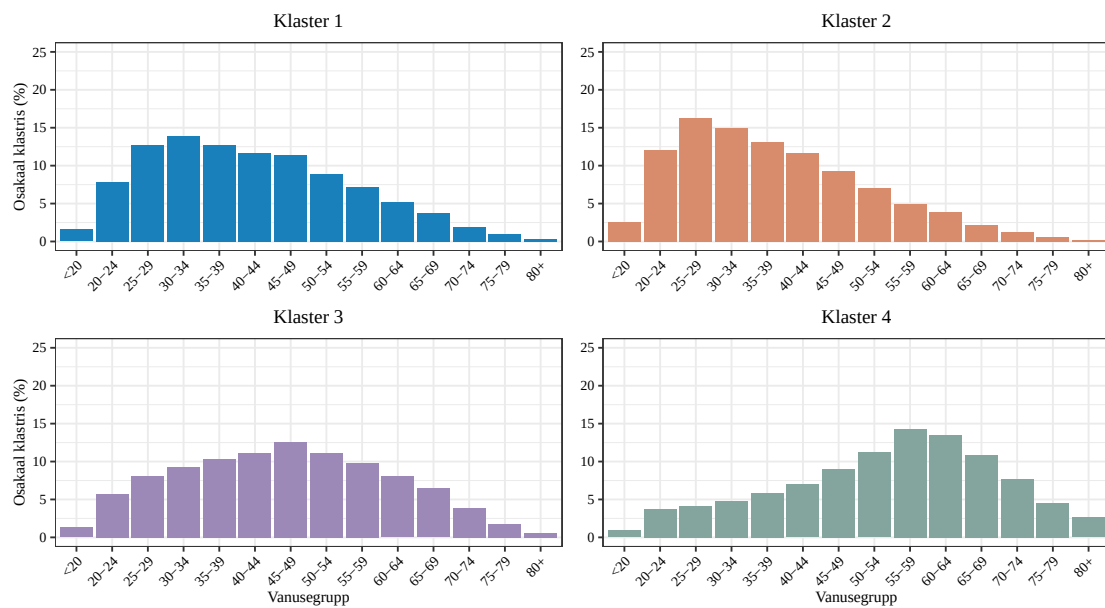


Joonis L1: K-keskmiste klasteranalüüsi jaoks optimaalsete klastrite arvu valimiseks rakendatud küünarnuki meetod $k \in [2,8]$ korral. $\sum WSS$ (*sum of within cluster sum of squares*) on klastrisiseste hajuvuste summa ehk $\sum_i^k W(C_i)$ ning optimaalseks klastritearvuks valiti mõlema kohordi korral $k = 4$.

Lisa 3: Vanuseline jaotus klastrites



Joonis L2: 2009.–2010. aastate kohordi klastrite vanuseline jaotus



Joonis L3: 2019.–2020. aastate kohordi klastrite vanuseline jaotus

Lisa 4: Diagnoosigruppidesse kuuluvad diagnoosid

Tabel L2: Diagnoosigrupid ja neisse kuuluvad diagnoosid 2003. aasta Eesti tervisekaotuse uuringust

Grupp	RHK-10 kood ja diagnoos
Unipolaarne depressioon	F32 depressioon
	F33 korduv depressioon
Ärevushäired	F40 foobiad
	F41 muud ärevushäired
Hüpertensioon	I10 kõrgvererõhktõbi
	I11 südamekahjustusega hüpertooniatõbi
	I12 neerukahjustusega hüpertooniatõbi
	I13 südame- ja neerukahjustusega hüpertooniatõbi
	I15 sekundaarne hüpertensioon
Südame isheemiatõbi	I20 rinnaahistus
	I21 äge müokardiinfarkt
	I22 korduv müokardiinfarkt
	I23 ägeda müokardiinfarkti mõned lähtüsisistused
	I24 südame ägeda isheemiatõve muud vormid
	I25 krooniline südame isheemiatõbi
Südame rütmihäired	I44 atrioventrikulaar- ja Hisi kimbu vasaksääre blokaad
	I45 südame erutusjuhte muud häired
	I47 paroksüsmaalne tahhükardia
	I48 kodade virvendus ja laperdus
	I49 muud südamerütmihäired e arütmiaid
Südamepuudulikkus	I50 südamepuudulikkus

Lisa 5: Dimensioonide keskmiste hinnangute mitmene võrdlemine

Dimensioonide keskmiste võrdlemiseks tehti vähimruutude keskmiste paariviisilisi t-teste ning kasutati Bonferroni korrektsiooniga p-väärtusi, st kõik esialgsest t-testist saadud p-väärtused korrutati vanusegruppide arvuga ehk 13-ga, et hoida kontrolli all I liiki vea tegemise tõenäosust. Kui korrigeeritud p-väärtus tuli suurem või võrdne ühega, märgiti p-väärtus võrdseks ühega.

Tabel L3: Liikuvuse dimensiooni keskmiste võrdlemine Bonferroni korrektsiooniga

Vanusegrupp	2009–2010	2019–2020	p-väärtus
<20	1,02 (1,01, 1,03)	1,05 (1,03, 1,06)	0,823
20–24	1,03 (1,02, 1,04)	1,04 (1,03, 1,04)	1,000
25–29	1,03 (1,02, 1,03)	1,03 (1,03, 1,03)	1,000
30–34	1,04 (1,03, 1,05)	1,04 (1,04, 1,04)	1,000
35–39	1,04 (1,03, 1,05)	1,05 (1,05, 1,05)	1,000
40–44	1,08 (1,07, 1,10)	1,06 (1,06, 1,07)	0,325
45–49	1,12 (1,11, 1,14)	1,09 (1,09, 1,10)	0,006
50–54	1,17 (1,15, 1,19)	1,14 (1,14, 1,15)	0,014
55–59	1,25 (1,23, 1,27)	1,20 (1,19, 1,21)	< 0,001
60–64	1,31 (1,28, 1,33)	1,26 (1,25, 1,27)	< 0,001
65–69	1,42 (1,39, 1,45)	1,31 (1,29, 1,32)	< 0,001
70–74	1,51 (1,48, 1,54)	1,36 (1,34, 1,38)	< 0,001
75–79	1,63 (1,59, 1,67)	1,43 (1,41, 1,46)	< 0,001

Tabel L4: Enesehoolitsuse dimensiooni keskmiste võrdlemine Bonferroni korrektsiooniga

Vanusegrupp	2009–2010	2019–2020	p-väärtus
<20	1,01 (1, 1,01)	1,01 (1, 1,01)	1,000
20–24	1 (1, 1,01)	1,01 (1,01, 1,01)	1,000
25–29	1 (1, 1,01)	1,01 (1, 1,01)	1,000
30–34	1,01 (1, 1,02)	1,01 (1,01, 1,01)	1,000
35–39	1,01 (1,01, 1,02)	1,01 (1,01, 1,01)	1,000
40–44	1,02 (1,01, 1,03)	1,02 (1,01, 1,02)	1,000
45–49	1,03 (1,02, 1,04)	1,02 (1,02, 1,02)	0,117
50–54	1,05 (1,04, 1,06)	1,03 (1,03, 1,04)	< 0,001
55–59	1,07 (1,06, 1,09)	1,05 (1,05, 1,06)	< 0,001
60–64	1,11 (1,09, 1,13)	1,07 (1,06, 1,07)	< 0,001
65–69	1,16 (1,14, 1,19)	1,06 (1,05, 1,07)	< 0,001
70–74	1,18 (1,16, 1,21)	1,07 (1,06, 1,08)	< 0,001
75–79	1,24 (1,21, 1,27)	1,08 (1,07, 1,1)	< 0,001

Tabel L5: Igapäevatoimetuste dimensiooni keskmiste võrdlemine Bonferroni korrektsiooniga

Vanusegrupp	2009–2010	2019–2020	p-väärtus
< 20	1,06 (1,04, 1,07)	1,11 (1,09, 1,13)	0,010
20–24	1,05 (1,04, 1,06)	1,10 (1,09, 1,11)	< 0,001
25–29	1,05 (1,04, 1,06)	1,07 (1,06, 1,08)	0,369
30–34	1,08 (1,06, 1,09)	1,07 (1,06, 1,07)	1,000
35–39	1,08 (1,07, 1,09)	1,08 (1,07, 1,08)	1,000
40–44	1,12 (1,10, 1,14)	1,09 (1,08, 1,09)	0,025
45–49	1,18 (1,16, 1,20)	1,11 (1,10, 1,12)	< 0,001
50–54	1,23 (1,21, 1,25)	1,16 (1,15, 1,16)	< 0,001
55–59	1,33 (1,31, 1,35)	1,22 (1,21, 1,23)	< 0,001
60–64	1,38 (1,36, 1,41)	1,26 (1,25, 1,28)	< 0,001
65–69	1,50 (1,47, 1,54)	1,28 (1,27, 1,30)	< 0,001
70–74	1,58 (1,55, 1,61)	1,33 (1,31, 1,35)	< 0,001
75–79	1,71 (1,66, 1,75)	1,39 (1,36, 1,42)	< 0,001

Tabel L6: Valu dimensiooni keskmiste võrdlemine Bonferroni korrektsiooniga

Vanusegrupp	2009–2010	2019–2020	p-väärtus
< 20	1,20 (1,18, 1,23)	1,31 (1,28, 1,34)	< 0,001
20–24	1,20 (1,19, 1,22)	1,29 (1,27, 1,30)	< 0,001
25–29	1,22 (1,20, 1,24)	1,26 (1,25, 1,27)	0,110
30–34	1,28 (1,26, 1,31)	1,27 (1,26, 1,28)	1,000
35–39	1,32 (1,30, 1,34)	1,33 (1,31, 1,34)	1,000
40–44	1,36 (1,34, 1,39)	1,37 (1,36, 1,38)	1,000
45–49	1,42 (1,40, 1,44)	1,42 (1,41, 1,43)	1,000
50–54	1,50 (1,48, 1,53)	1,48 (1,47, 1,49)	1,000
55–59	1,57 (1,54, 1,59)	1,55 (1,54, 1,56)	1,000
60–64	1,60 (1,58, 1,63)	1,60 (1,58, 1,61)	1,000
65–69	1,67 (1,64, 1,70)	1,62 (1,61, 1,64)	0,084
70–74	1,71 (1,68, 1,74)	1,67 (1,64, 1,69)	0,178
75–79	1,74 (1,70, 1,77)	1,68 (1,65, 1,71)	0,180

Tabel L7: Ärevuse dimensiooni keskmiste võrdlemine Bonferroni korrigeerimisega

Vanusegrupp	2009–2010	2019–2020	p-väärtus
< 20	1,30 (1,27, 1,33)	1,50 (1,47, 1,54)	< 0,001
20–24	1,32 (1,30, 1,34)	1,47 (1,46, 1,49)	< 0,001
25–29	1,33 (1,31, 1,35)	1,41 (1,40, 1,43)	< 0,001
30–34	1,38 (1,35, 1,40)	1,38 (1,37, 1,39)	1,000
35–39	1,39 (1,37, 1,42)	1,39 (1,38, 1,40)	1,000
40–44	1,40 (1,37, 1,42)	1,39 (1,38, 1,40)	1,000
45–49	1,42 (1,40, 1,45)	1,36 (1,35, 1,37)	< 0,001
50–54	1,45 (1,42, 1,47)	1,38 (1,36, 1,39)	< 0,001
55–59	1,48 (1,46, 1,51)	1,38 (1,36, 1,39)	< 0,001
60–64	1,47 (1,44, 1,49)	1,39 (1,37, 1,40)	< 0,001
65–69	1,47 (1,44, 1,50)	1,37 (1,35, 1,38)	< 0,001
70–74	1,51 (1,48, 1,54)	1,38 (1,36, 1,40)	< 0,001
75–79	1,50 (1,46, 1,54)	1,36 (1,33, 1,39)	< 0,001

Lisa 6: Euroopa *value set*

Geenidonorite EuroQoli indeksite väärtuste arvutamiseks kasutati Euroopa *value set*'i, sest Eestis pole selle rahvastikku kirjeldavat *value set*'i veel loodud.

Tabel L8: W. Greineri 2003. aastal loodud Euroopa *value set*

Dimensioon	Hinnang	Kaal
Liikuvus	1	0
	2	0,0659
	3	0,1829
Enesehoolitsus	1	0
	2	0,1173
	3	0,1559
Igapäevaste tegevustega toimetulek	1	0
	2	0,0264
	3	0,0860
Valu ja ebamugavus	1	0
	2	0,0930
	3	0,1637
Ärevus ja depressioon	1	0
	2	0,0891
	3	0,1290
Vähemalt üks hinnang on suurem kui 1		0,2288
Vähemalt üks hinnang on suurem kui 2		0,1279

Lisa 7: Indeksi väärtuse dispersioonanalüüsi mudelid soo ja kohordi järgi

Nii aastate 2009–2010 kui ka 2019–2020 korral tehti eraldi dispersioonanalüüsi mudelid meeste ja naiste jaoks, et uurida argumenttunnuste seost indeksi väärtusega. Argumenttunnusteks mudelis olid vanusegrupp (alates 25-aastastest kuni 79-aastasteni) ning haridusaste.

Tabel L9: Dispersioonanalüüsi tulemused 2009.–2010. aastate kohordi naiste korral

Faktor	Vabadusastmete arv	Ruutude keskmine	F-statistik	p-väärtus
Vanusegrupp	10	5,135	163,246	< 0,001
Haridusaste	2	5,433	172,735	< 0,001
Vanusegrupp × haridusaste	20	0,065	2,069	0,003
Jääk	11061	0,031		

Tabel L10: Dispersioonanalüüsi tulemused 2009.–2010. aastate kohordi meeste korral

Faktor	Vabadusastmete arv	Ruutude keskmine	F-statistik	p-väärtus
Vanusegrupp	10	2,105	61,437	< 0,001
Haridusaste	2	3,194	93,227	< 0,001
Vanusegrupp × haridusaste	20	0,084	2,444	< 0,001
Jääk	6011	0,034		

Tabel L11: Dispersioonanalüüsi tulemused 2019.–2020. aastate kohordi naiste korral

Faktor	Vabadusastmete arv	Ruutude keskmine	F-statistik	p-väärtus
Vanusegrupp	10	5,756	206,583	< 0,001
Haridusaste	2	11,065	397,146	< 0,001
Vanusegrupp × haridusaste	20	0,053	1,914	0,008
Jääk	47231	0,028		

Tabel L12: Dispersioonanalüüsi tulemused 2019.–2020. aastate kohordi meeste korral

Faktor	Vabadusastmete arv	Ruutude keskmine	F-statistik	p-väärtus
Vanusegrupp	10	2,423	91,051	< 0,001
Haridusaste	2	5,895	221,502	< 0,001
Vanusegrupp × haridusaste	20	0,039	1,464	0,082
Jääk	21994	0,027		

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Lisett Klement,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „TÜ Eesti geenivaramu geenidonorite tervisega seotud elukvaliteet“, mille juhendajad on Krista Fischer ja Taavi Lai, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Lisett Klement

12.05.2026