

Tartu Ülikool
Sotsiaalteaduste valdkond
Psühholoogia instituut

Raul Jack Barrow

Geenivaramu isiksuse andmestiku esinduslikkuse analüüs

Uurimistöo

Juhendaja: Uku Vainik, *PhD*

Läbiv pealkiri: Geenivaramu andmestiku esinduslikkuse analüüs

Tartu 2025

Geenivaramu isiksuse andmestiku esinduslikkuse analüüs

Lühikokkuvõte

Esinduslikkus on igas uuringus vajalik, kuid sageli töötavad psühholoogid mitte-esinduslike andmestikega. Mitte-esinduslikkuse ulatus ja mõju isiksuseuuringutes on Eestis suuresti teadmata. Võrdlesime kahte kõrgema osaluskalde riskiga andmestikku (EstBB ja RVTU) madalama osaluskalde riskiga andmestikega (RAHA ja SALK). Me kasutasime Student t -testi, et leida andmestike vahelisi olulisi erinevusi isiksust mõõtvate väidete keskmistes ning kontrollisime väidete keskmiste erinevuste kattuvust samade väidete demograafiliste korrelatsioonide erinevustega. Vähemalt pooled väidete keskmistest erinesid oluliselt kõikide andmestike vahel, mitte ainult kõrgema vs madalama riskiga andmestike vahel. Väidete korrelatsioonid demograafiliste näitajatega erinesid väheses koguses ning keskmiste erinevusega väited polnud samad, mis korrelatsiooni erinevustega väited. Meie leidsime, et andmete vahel on väidetel keskmiste erinevused, kuid nende erinevuste olemasolu ei lange kokku erinevustega demograafilistes korrelatsioonides.

Võtmesõnad: Isiksus, Eesti Geenivaramu, Suur viisik

An Analysis of Representativeness in the Personality Dataset of the Estonian Biobank**Abstract**

Representativeness is an integral part of any research, but psychologists often work with non-representative datasets. The extent of non-representativeness and its effect in personality research in Estonia is largely unknown. We compared two datasets with a higher risk of participation bias (EstBB and RVTU) to two datasets with a lower risk of participation bias (RAHA and SALK). We used the Student's *t*-test to find whether the means of statements meant to measure personality were significantly different between datasets and checked whether those differences aligned with differences in demographical correlations. At least half of the statements' means differed significantly between all datasets, as opposed to just higher vs lower risk datasets. The statement's correlations with demographical variables differed in a small amount of cases and statements with differences in means were different than statements with differences in correlations. We found that there were differences in means between datasets, but the existence of differences did not align with differences in demographical correlations.

Keywords: Personality, Estonian Biobank, Big Five

Viimaste aastate jooksul on palju uuritud valimite esinduslikkust. Uuringute läbiviimisel on esinduslikkusel oluline roll, sest sellest võib olla suuresti mõjutatud uuringust tehtavad järeldused (Schoeler jt., 2023; Van Alten jt. 2022). Suuremad mitmekümnete tuhandete osalejatega uuringud, jätavad turvalise mulje, kuna nii suure valimi puhul ootaks, et andmed on täiesti üldistatavad ka üldpopulatsioonile ja valimi suurusega seoses on esindatud täiesti mitmekülgne lõik populatsioonist. Sõltumata valimi suurusest, kui tegu on vabatahtlike osalejatega, võib aga tekkida andmeid oluliselt mõjutav probleem – osaluskalle. Antud töö eesmärk on uurida Eesti Geenivaramu valimi isiksuseomaduste esinduslikkust ja mõjutatust osaluskaldest.

Osaluskalle tekib, kui vastajate teatud omadus suurendab tõenäosust, et nad uuringus osalevad. Ka suur valim võib esindada ainult piiratud hulka üldpopulatsioonist. Sellega seoses võivad tekkida isegi valesuunalised seosed (Van Alten jt., 2022) või riskifaktorite mõju võib olla tegelikus populatsioonis oluliselt erineva suurusega, kui osaluskaldega valim näitaks.

Osaluskallet on uuritud näiteks Norras, võrreldes kutsutud valimi ning kutsele vastanud inimeste demograafilisi ja terviseandmeid (Søgaard jt., 2004). Võrreldes 40000 kutsutuga 19000 osaleja seas oli vähenenud representatsioon noortest, meestest, lahutatud või mitteabiellunudest, immigrantidest, kesklinna elanikest, madala majandusjõuga inimestest ning puudetoetusega inimestest. Uurijate sõnul andmete üldistatavus ei olnud mõjutatud demograafilise esindatuse puudujääkidest. Täpsemalt vaid puudetoetusega inimeste madal esinemine võis levimuse kirjeldamisel probleemiks olla ning muud demograafilised tegurid ei avaldanud olulist mõju levimuse kirjeldusele. Søgaard jt. (2004) sõnul ei pruugi osaluskalle levimuse kirjeldamisel suuremaks probleemiks olla, kuid esines nende uuringus oluliste demograafiliste erinevuste kujul.

Vastukäivalt leidsid Schoeler jt. (2023), et inglismaa geenivaramu (UK Biobank) esinduslikkus oli puudulik ning osaluskalle avaldas olulist mõju levimuse kirjeldamisel. Schoeler jt. (2023) analüüsisid üksiknukleotiidne polümorfismide (*single nucleotide polymorphism, SNP-de*) seoseid ja seoste efektsuuruseid riskifaktorite levimusele. Võrreldes esindusliku valimiga olid geenivaramu osalejad tervemad, vanemad ja leibkonnati suurema sissetulekuga. Kui nad hiljem kaalusid SNP-de efekte esindusliku valimist tulenenud erinevuste vähendamiseks, olid SNP-de efektide suuruste suunad samad, kuid oluliselt suurendatud või vähendatud. Levimus oli

selle tõttu liialdatud vähi, üksilduse, haridustaseme ja reaktsiooniaja puhul ning alahinnatud masenduse/ärevuse, kohvi tarbimise ja suitsetajate levimuse puhul. Järelikult osaluskalle saab levimuse kirjeldamist mõjutada.

On uuritud ka põhjusi miks osaluskalle võib tekkida. Üheks pakutavaks põhjuseks on olnud isiksuste erinevuste mõju osalusele ning isiksusejoontest tingitud suurenenud või vähenenud tõenäosus vabatahtlikult uuringus osaleda. Erinevate suure viisiku ja psühhopatoloogiliste tunnuste mõju osalusele on mitmekordselt uuritud ning on leitud seoseid teatud omaduste ning osaluse või uuringu kestel välja langemise vahel (Bosnjak jt., 2013; Marcus & Schütz, 2005; Pagan jt., 2006).

Bosnjak jt. (2013) viisid läbi uuringu Saksamaal, kus kutsuti ALLBUS küsitlusele vastajaid osalema netifoorumis. ALLBUS küsitlus oli randomiseeritud, piirkonniti grupeeritud ning seetõttu enamasti esinduslik. Osalejate välja langemist valimist vaadati mitmes järgus ning võrreldi Allbusi algse valimiga nii sotsiaaldemograafiliselt, kui ka isiksusejoontega, kasutades suure viisiku mudelit. Kutsutud langesid algselt välja, kui neil ei olnud internetti – antud voorust edasipääsejad, ehk netikasutajad olid rohkem avatud, vähem meelegindlad, nooremad, kõrgemalt haritud ja tavaliselt mehed. Järgnevalt langesid välja inimesed, kes ei soovinud foorumis osaleda – kes alles jäid olid vähem meelegindlad, nooremad ja kõrgemalt haritud. Sõltumata algsest osalussoovist, ei vastanud mõlemale eelnevale küsitlusele osad osalejad. Lõplikud vastajad olid võrreldes algse ALLBUS küsitlusele vastanutega keskmiselt nooremad, haritumad, rohkem avatud ja vähem meelegindlad. Nii demograafiliste tunnuste, kui isiksusejoonte osas oli oluline vahe. Foorumi valim oli seetõttu vähem esinduslik, kuigi osalejad olid võetud samast populatsioonist. Värbamise käigus ja uuringu jooksul langes uuringust välja suur ja varieeruvust pakkuv osa populatsioonist.

Kui eelnevas uuringus oli isiksuste hinnangute andmed eneseküsimustike baasil, uurisid Pagan jt. (2006) nii eneseküsimustike, kui ka kõrvaliste vaatlejate hinnangute abil katseisikute isiksusejooni. Osalejad hindasid psühhopatoloogilisi tunnuseid endas ning teistel tudengitel enda ühika korrusel. Mitte-osalejad, kes algselt lubasid osaleda, kuid ei vastanud lõpuks küsimustikele, olid võrreldes osalejatega nartsissistlikumad ning vähem kehtestavad (ümbritsevate inimeste hinnangul, kuna nad ise küsimustikke ei täitnud). Osalejad olid enda ja teiste tudengite hinnangul rohkem obsessiiv-kompulsiivsete, histriooniliste, teistest-sõltuvate ja

ennast-ohverdavate tunnustega. Antud uuringus tuli taas välja osaluskaldega kaasnevad oluliselt suured erinevused osalevas valimis ja kutsutud isikute vahel.

Inimeste hindamisel võib esineda teatud mõjusid inimeste tuttavusest, näiteks eelmises uuringus, esines sellest tulenevalt andmete kalle, millega püüti arvestada. On ka uuritud võõraste isikute isiksuste hindamise läbi valimite esinduslikkust. Marcus & Schütz (2005) uurisid uuringu eri etappidest välja langemise võimalike seoseid isiksusejoontega. Valim koosnes isiklike veebilehtede omanikest (kellest vastajad vastasid ka enesekohasetele isiksust hindavatele küsimustikele) ning vaatlejad olid vabatahtlikud, kes hindasid veebilehtede kaudu omanike isiksusi. Veebilehe omanikest vastajad olid mittevastanutest vaatlejate hinnangul koostöövalmimad ja kogemusele rohkem avatud. Demograafilistele küsimustele meelega vastamata jätnud isikud olid vähem ekstravertsed. Jätkuintervjuust huvitatud vastajad olid rohkem ekstravertsed, avatud kogemusele ning vähem meelekindlad (nii enesekohaste, kui vaatluse põhjal). Vastupidi Pagan jt. (2006) uuringule leiti, et nartsissistlikud isiksusejooned just tõstsid tõenäosust jätkuintervjuus osaleda. Vahe võis tuleneda küsimustiku täitmise erinevusest näost näkku intervjuuga.

Van Alten jt. (2022) püüdsid leida viisi osaluskalde vähendamiseks. Inglismaa vabatahtliku osalusega geenivaramut (UKB) võrreldi randomiseeritud Inglismaa rahvaloendusega. Võrreldi UKB ja rahvaloenduse kohandatud valimit. Rahvaloenduse valimit kohandati nii, et nende sotsiaaldemograafilised andmed ühtiksid UKB-ga – võrdsustati osalejate piirkonnad, vanusepiirid ja muud sotsiaaldemograafilised tunnused. van Alten jt. (2022) leidsid esinduslikkuse osas probleeme geenivaramuga – osalejad olid UKB esindusliku valimiga võrreldes vanemad, tervemad, kõrgema hariduse ja kõrgema sotsiaalmajandusliku staatusega. UKB andmetel oli veel ebaloogilisi nihkeid reaalsusega – esinduslikus valimis oli kaks korda tõenäolisem, et osalejad väidaksid, et neil on halb tervis, kuigi UKB valim oli rahvaloenduse valimist keskmiselt vanem, mis räägib vastu vanuse kasvades esinevale füüsilisele allakäigule. Esinduslikust valimist 73.9% omasid maja, aga UKBs 89.9%.

Kohandatud rahvaloenduse andmeid kasutades koostati mudel korrigeerimaks UKB esinduslikkust (Van Alten jt., 2022). Mudeli kaalude tagajärjel hakkas UKB info reaalsusega ühtima – enne korrigeerimiseks kasutatud statistilist kaalumist oli enne 1950. aastat sündinud inimesed tõenäoliselt mehed, kuigi faktuaalselt on vastupidi; 50. aastal sündimine oli nõrgalt

seotud hea tervisega, kuigi on vastupidi ja veel sarnaste ebaloogiliste andmete parandusi. Kaalumise tegi Inglismaa geenivaramust tehtud järeldused loogilistemaks ja tegelikele demograafiliste andmetega klappivaks. Van Alten jt. (2022) sõnul vähendas nende meetod suurel hulgal osaluskaldest tekitatud andmete kallet.

Erinevate andmestike erinevused võivad ka olla tingitud erinevate andmekogumisfirmade kasutusest. Kui erinevad firmad haldavad endas erinevat osa populatsioonist, kas juhuslikult või näiteks värbamise või andmete kogumise viisi osas, võib esineda koostatud valimis nihkeid. Mölder (2021) leidis, et Eesti kodanike erakondade toetamise protsent on süstemaatiliselt erinev üle Eesti turuuringute firmade. Sama on leidnud ka Jüristo (2021), täheldades erinevusi Kantar/EMOR, Norstat ja Turu-Uuringute AS-i erakondade toetuse uuringute andmetes. Üheks pakutud erinevuste mõjuriks on, et kuigi valim on ehitatud esinduslikel andmetel, on osalejad andnud eelnevalt nõusoleku uuringu kutseid saada ning värbamise ja andmekogumise viisid erinevad uuringufirmade vahel (Jüristo, 2021; Mölder, 2021).

Geenivaramu hõlmab endas 20% Eesti täiskasvanutest. Avastades geenivaramu osaluskallet ja selle võimaliku mõju, võib paremini aru saada, kas geenivaramust on realistlik luua seoseid ka ülejäänud Eesti populatsiooniga. Leides, esinduslikkuse kalde võimalikku mõju, oleks võimalik teha järelduste ja levimuste kirjeldamise üldpopulatsiooni kajastavamaks.

Kasutades esinduslikkuse põhimõtetel läbiviidud uuringuid, saan võrrelda, kas geenivaramul esineb osaluskalle teatud isiksusjooni mõõtvate väidete puhul. Olemasolevaid andmeid kasutades viin läbi statistilise analüüsi esinduslikest valimitest ja ka nende ja geenivaramu valimi vahel. Juhul kui geenivaramus esineb osaluskalle saan uurida ka samade valimite abil, kas korrelatsioonides esineb erinevusi, ehk kas osaluskalle mõjutab geenivaramu isiksuste korrelatsioonid soo, hariduse ja vanusega.

Millistes geenivaramu suure viisiku ja/või isiksust kirjeldavates väidetes esineb osaluskalle?

Kuidas geenivaramu isiksusjoonte esinduslikkuse kalle langeb kokku samade väidete kalletega seostes soo, vanuse ja haridustasemega?

Meetod

Andmed

Estonian Biobank Personality Study (EstBB PS)

Eesti geenivaramu projekti raames küsitleti osalejaid ajavahemikul november 2021 kuni aprill 2022 (Milani jt., 2025). Geenivaramu osalejatele saadeti kutse ja paluti vabatahtlikult osaleda antud katses. Vastanuid uuriti isiksuse omaduste, sotsiaalmajanduslike ja demograafiliste andmete osas. Isiksuste mõõtmiseks kasutati 100NP küsimustikku. 100NP sisaldab 198 küsimust, mis kajastab suurt viisikut, HEXACO-d ning muid omadusi, juhul kui spetsiifiliseks uurimiseks on vajadust sellele.

Antud valim koosneb geenivaramust eraldi küsitletud kohortist, mitte tervest geenivaramust (Arumäe jt., 2025). Valimis on 77400 inimest, kellest 54548 (70,5%) olid naised. Vastajate keskmine vanus oli 47.5 (SD = 14.6, 18-102). Vastajatest oli 93.7% eestlased, 4.7% venelased ja 1,6% muu. Enamus vastajatest oli kõrgharidusega (58%), 38% kutsekesk- või keskharidusega, 3.2% põhihariduse või madalamaga.

RAHA

„Financial well-being & personality study” uuringut juhtis Leonore Riitsalu, uuringu viis läbi Rein Urmas Murakas. Valim on koostatud niimoodi, et ta sisaldaks esinduslikku representatsiooni Eesti populatsioonist eesti rahvastikustatistika sotsiaaldemograafilistel andmetel. Antud valim peaks seetõttu esindama aktuaalselt läbilõiget eesti rahvastikust ning saame seda kasutada võrdlustel geenivaramu valimiga.

RAHA valim koosneb 1007st isikust kellest 51% on naised 48.5% mehed ja 0.5% muu soomääratlusega. Valim on jaotatud vanuserühmadesse, kellest suurim on 40-64 aastased (47.4%), 25-39 aastased moodustavad valimist 32.6%, 65-74 aastased 14.4% ja 18-24 aastased 5.6%. Osalejatest 12.8% olid Põhihariduse või vähemaga, 55.9% keskharidusega ja 32.8% kõrgharidusega.

Andmed olid kogutud netiküsitluse teel Limesurvey kaudu ja edastatud läbi Norstati. Küsimustik sisaldas endas küsimusi rahalise heaolu, isiksusjoonte ja sotsiaaldemograafiliste andmete kohta. Isiksuse kohta käivaid küsimusi oli 22, mis oli võetud 198st väitest isiksuse

hindamiseks, millest kõik 198 on kasutusel geenivaramu uuringul (NP100 küsimuste kujul). Küsimused on suure viisiku isiksusejoonte hindamiseks kasutatavad.

SALK

SALK uuringu viis läbi Sihtasutus Liberaalne Kodanlik Tarmo Jüristo juhtimisel. Isiksuse osa uuringust nõustas Anu Realo. Valimisse on võetud osalejaid, et eesti populatsiooni esindada. Uurijad kasutasid ristkvoote rahvus-haridus-vanus-sugu-regioonid ning neid on kaalutud üldkogumile. Valim koosneb 1000st vastajast. Nendest olid 54.7% naised ja osalejatest 78.9% täitsid paneeli eesti keeles ja 21.1% vene keeles. Valimis osalejate keskmine vanus oli 48.57 (SD = 17.82, 18-96). Rahvuselt olid osalejatest 81.7% märkinud, et nad on eestlased ja 18.3% muust rahvusest. Haridustasemeks märkisid 52.4% osalejatest keskhariduse või kutsekeskhariduse, 33.7% kõrghariduse ja 13.9% alghariduse, põhihariduse või kutsehariduse põhikooli baasil.

Andmed koguti veebipõhiselt, värvates telefoni teel vabatahtlikke osalejaid veebipaneeli. Isiksuse omadusi mõõdeti antud uuringus kasutades NP-100 küsimuste kogumit. Uurijad kasutasid piiratud arvu antud kogumiku küsimustest, mis kalduksid mõõtma vaid ühte kindlat isiksuse omadust ja ei ristuks teiste isiksuste omadustega. Küsimused valiti samuti nii, et oleksid nii eesti kui vene keeles usaldusväärsed. Kokku kasutati 20 küsimust, 4 küsimust iga suure viisiku omaduse kohta.

RVTU

„Rahvusliku vaimse tervise uuringu“ (RVTU) uurijad kasutasid stratifitseeritud süstemaatilist juhusliku valimist, et Eesti Rahvastikuregistrist valida välja 20000 osalejat (1. jaanuar, 2020 olemasolevate andmete baasil). Uuring toimus kolmes laines, osaliselt varieeruvate küsimustega. Osalejaid värvati kas veebi- või postiteel ning osalemine oli vabatahtlik. Võtsime kasutusele vaid kolmanda laine osalejad, selle tõttu, et ainult kolmandas laines kasutati NP-100 küsimustiku kogumi väiteid. Enne kolmandat lainet uuendati valimi info, eemaldades valimist surnud või väljarännanud osalejad. Kolmanda laine andmed koguti kahe kuu jooksul alates 3. jaanuarist, 2022. Kolmandas laines oli 4744 osalejat, kuid pärast andmepuhastust jäi alles 4495 osalejat, kes vastasid 22le NP-100 väitele (4 neurootilisust, ekstravertsust, meelekindlust ja 5 avatust kogemusele ning koostöövalmidust kajastavat väidet. Kolmanda laine osalejatest olid 61.8% naised. Kõrgharidusega osalejaid oli 31.7% ning osalejate keskmine vanus oli 55.72 (SD

= 18.34, 18-97). Osalemise ajendamiseks loositi kolmandal lainel 50 toidupoe kinkekaarti väärtuses 30 eurot ning kõigides kolmes laines osalenute vahel 20 toidupoe kinkekaarti väärtuses 100 eurot.

Esinduslikkus

Jaotasime valimid esinduslikeks ja mitte-esinduslikeks. Esinduslikeks valimiteks olid RAHA ja SALK valimid, kus oli kasutatud eelpool mainitud osalejate värbamise strateegiaid, et koostada esindusliku valimi. Kategoriseerisime EstBB mitte-esinduslikuks selle tõttu, et osalemine oli vabatahtlik. Kategoriseerisime RVTU valimi mitte-esinduslikuks tingituna uuringu lainete vahel osade vabatahtlike osalejate välja langemisest.

Statistiline analüüs

Statistilisi analüüse viidi läbi programmides JASP, JAMOVI ja Excel.

Nelja andmestiku vahel ühtis 44 isiksust hindavat väidet, millest kaheksa hindas neurootilisust, seitse ekstraversust, kaheksas avatust kogemusele, 13 abivalmidust ning üheksa meelekindlust. Edaspidi ilmuvad väited on teksti sees mainitud NP-100 väitenimetuse lühendatud versioonina. Väidetele on tähemärk vastavast isiksusjoone grupist (neurootilisus – N, ekstraversus – E, avatus kogemusele – O, abivalmidus – A, meelekindlus – C), numbriga vastavalt asetusele oma isiksusjoone grupis NP-100 küsimustikus ning juhul kui küsimus oli pööratud, esineb lühend ka koos „R“ tähega (andmestike ühised väited on koos väitesisuga olemas Tabelis 1). Näiteks väide Neuroticism08 oleks N08 või pööratud kujul N08R. ESTBB andmestikus esinesid kõik 44 väidet, RAHA andmestikus 22, SALK andmestikus 28 ja RVTU andmestikus 19. Väidetele sai vastata 1-6 Likert skaalal: 1 – “Täiesti vale”, 2 - “Enamasti vale”, 3 – “Pigem vale”, 4 – “Pigem õige”, 5 “Enamasti õige”, 6 “Täiesti õige”. Algselt leidsime kõikide andmestike vastavate väidete keskmised ja standardhälbed. EstBB, RAHA ja SALK andmestike keskmised ja standardhälbed arvutasime Excelis/LibreOffices ning RVTU uuringu keskmised ja standardhälbed olid meie jaoks ette arvutatud.

Tabel 1*Andmestike vahel ühised väited*

Väitekode	Lühend	Väide
neuroticism		
neuroticism03R	N03R	"Muretsen harva."
neuroticism08	N08	"Vihastan kergesti."
neuroticism09	N09	"Olen tihti õnnetu."
neuroticism10	N10	"Mu meeleolu muutub tihti."
neuroticism14	N14	"Muretsen, mida teised minust mõtlevad."
neuroticism19	N19	"Satun kergesti stressi."
neuroticism25	N25	"Tunnen tihti, et olukord on täiesti lootusetu."
neuroticism36	N36	"Hakkan kergesti nutma."
extraversion		
extraversion03	E03	"Mulle meeldivad seltskondlikud üritused."
extraversion08	E08	"Leian kergesti sõpru."
extraversion17	E17	"Olen tavaliselt aktiivne ja energiast pakatav."
extraversion19	E19	"Armastan ohtlikke olukordi."
extraversion21	E21	"Armastan põnevust."
extraversion26	E26	"Mul on tihti lõbus."
extraversion28	E28	"Naeratan tihti."
openness		
openness09R	O09R	"Väldin filosoofilisi arutelusid."
openness11	O11	"Mulle meeldib lugeda."
openness12	O12	"Armastan uusi asju õppida."
openness17	O17	"Mulle meeldib nuputada üha uusi viise, kuidas asju teha."
openness18	O18	"Pean saama end loominguliselt väljendada."
openness20	O20	"Mulle meeldib keerulisi probleeme lahendada."
openness28	O28	"Mind huvitavad paljud asjad."
agreeableness		
agreeableness02	A02	"Usaldan teisi."

Väitekood	Lühend	Väide
agreeableness08R	A08R	"Minu jaoks on lihtne teisi manipuleerida."
agreeableness13R	A13R	"Kasutan teisi inimesi enda huvides ära."
agreeableness14R	A14R	"Tahan, et mind kardetaks."
agreeableness28	A28	"Oskan teisi lohutada."
agreeableness33R	A33R	"Maksan kätte, kui keegi mind solvab."
agreeableness35R	A35R	"Usun, et olen teistest parem."
agreeableness41R	A41R	"Arvan, et mul on alati õigus."
agreeableness42R	A42R	"Süüdistan teisi, kui midagi valesti läheb."
agreeableness43	A43	"Mul on lihtne vabandada, kui olen eksinud."
agreeableness49R	A49R	"Mulle meeldib väga, kui mind tunnustatakse."
agreeableness51R	A51R	"Imetlen, kui keegi eriti osava pettusega hakkama saab."
agreeableness58R	A58R	"Naudin teistele haiget tegemist."
conscientiousness		
conscientiousness05R	C05R	"Väga harva saan asjadega õigeaks ajaks valmis."
conscientiousness06	C06	"Töötan kõvasti."
conscientiousness09	C09	"Hoian oma asjad korras."
conscientiousness12	C12	"Pean oma lubadusi."
conscientiousness14	C14	"Hakkan kohe pihta, kui on vaja midagi ära teha."
conscientiousness26	C26	"Ma ei jäta töid pooleli."
conscientiousness42	C42	"Oskan raha kokku hoida."
conscientiousness43	C43	"Kulutan enam raha, kui peaksin."
conscientiousness44R	C44R	"Väldin kohustusi."

Osaluskalde tuvastamiseks kasutasime Student *t*-testi, millega leidsime väidete statistiliselt olulisi erinevusi. Selleks kasutasime Excelis valemeid, mis arvestasid keskmise, standardhälbe ja valimi suurusega. RAHA ja SALK uuringutel oli 6 ühist küsimust. Plaanisime võrrelda esimesena nende ühiste väidete keskmisi ja standardhälbeid ning kui väidetel oleksid esinenud sarnased keskmised, oleksime liitnud kokku nii RAHA kui SALK uuringu väited ühiseks esinduslikuks võrdluspunktiks. Selle tulemusena oleksime saanud 44 väidet, millel oleksid olemas nii EstBB ja RVTU uuringu kui ka esindusliku valimi hinnang. Selle tõttu, et

RAHA ja SALK valimid olid esinduslikult üles ehitatud, peaksid nende valimite ühiste väidete keskmised olema suhteliselt sarnased ehk ei tohiks olla statistiliselt oluliselt erinevad. Juhul kui EstBB või RVTU uuringu andmete ühised väited RAHA ja SALK valimitega erinesid statistiliselt oluliselt, kajastaks see osaluskalde esinemist EstBB ja RVTU uuringu andmestikes. Juhuslike tulemuste saamise riski vähendamiseks kasutasime olulisuse piirarvuna $p = 0.005$, $t = 2.81$. Kõik valimid olid vähemalt 1000st osalejast koosnevad.

Võimaliku osaluskalde mõju leidmiseks leidsime väidete korrelatsioonid demograafiaga - sugu, haridus ja vanuse. Seejärel leidsime, kas väite-demograafia korrelatsioonid olid erinevates valimites erinevad. Korrelatsioonide leidmisel kasutasime Pearsoni R selle tõttu, et andmed olid 1-6 skaalal, mida võib modelleerida kui pidevat tunnust (Rhemtulla jt., 2012; Robitzsch, 2020). Kasutasime Fisheri r to z transformatsiooni, et leida andmestike vahel korrelatsioonide olulisi erinevusi. Konverteerisime Pearsoni R korrelatsioonid z -skoorideks ning kasutasime erinevuse olulisuse hindamiseks valemit $z_{\text{observed}} = (z_1 - z_2) / \sqrt{(1 - N_1 - 3) + (1 / N_2 - 3)}$ (Statistic Solutions, 16. jaanuar, 2025). Korrelatsioonide oluline erinevus näitas andmestike tulemuste erinevust. Juhul kui samad väited olid nii keskmiste erinevusega kui ka korrelatsioonide erinevusega, näitas see osaluskalde võimalikku mõju uuringute tulemustele. Juhuslike tulemuste saamise riski vähendamiseks kasutasime siin ka olulisuse piirarvuna $p = 0.005$, kus $z = 2.576$.

Eetika

Kasutasime andmeid andmete omanike loal. SALK ja RAHA uuringud ei vajanud eetikakomitee nõusolekut kuna osalejad olid anonüümsed (Murakas, 2025). Me ei ole andmeid kasutanud viisil, mis võimaldaks osalejaid tuvastada ning andmed olid säilitatud uurimistöö autoril vaid uurimistöö kirjutamise ajal. Me ei analüüsinud RVTU andmeid - need analüüsid tegid meie eest R keskkonnas RVTU vastutav uurija Kenn Konstabel.

EstBB PS andmed olid kogutud Eesti Geenivaramu projekti “Geneetilise varieeruvuse mõju hindamine inimese isiksusele” raames, ajaperioodil november 2021 kuni märts 2022, mille läbiviimiseks taotleti 13.04.2020 eetikaluba 1.1-12/626 Eesti bioetika ja inimuuringute nõukogult. Uuring viidi läbi vastavalt Inimgeeniuringute seaduse nõuetele, mis hoomab muuhulgas nii geenidoonorite õigusi kui ka andmekaitse põhimõtteid. Geenivaramuga ühinev geenidoonor allkirjastab laia nõusoleku, mille kohaselt ta annab loa enda kohta kogutud andmeid

siduda teiste andmebasidega ning kasutada neid erinevates teadusuuringutes. Andmed säilitatakse Eesti Geenivaramu serverites, millele on ligipääs vaid autoriseeritud isikutel. Uurimistöö autoriga sõlmiti konfidentsiaalsusleping, mille järgselt võimaldati vaid uurimistöö eesmärgi saavutamiseks seotud pseudonüümiseeritud kujul andmetele ligipääs turvalises TÜ teadusarvutuste keskuse serveris väljastusloa nr3-10/GI/11571 alusel. Isikuandmeid uurijatele ei väljastata ning ühtegi konkreetset isikut ei ole võimalik kasutatud andmetega seostada.

Autori panus

Teadusartiklite analüüsimine, kokkuvõtmine, ümbersõnastamine ja uurimistöösse struktureerimine/mahutamine. Kasutatavate andmete analüüsimine ja kirjeldamine. Uurimistöö teema tutvustamine, kirjeldamine ja lahti seletamine. Uurimistöö vormistamine vastavalt APA nõuetele. RVTU andmed analüüsis Kenn Konstabel.

Tulemused

Andmete puhastus ja kirjeldav statistika

EstBB algsest valimist langes välja 166 vastajat. Antud vastajatel oli küsimustike vastustes puuduv informatsioon või küsimustikud lõpetamata. RAHA valimist langes välja kaks vastajat selle tõttu, et nad valisid sugu tunnuseks „muu soomääratlus“, mis ei esinenud teiste andmestike valikuvariantides. Hariduse tunnus oli andmestike vahel erinevate tasemete ja tasemete arvuga, nii et me jagasime kõikides andmestikes hariduse tunnuse kahte kategooriasse: kõrgharidus – kui osalejal esines rakenduslik- või kutsekõrgharidus, kõrghariduse põhiõpe (bakalareus) ja kõrghariduse kraadiõpe (magister, doktor); mitte-kõrgharidus – kui osalejal esines muu ehk madalam haridustase kõrghariduse kategooria variantidest. RAHA andmestikus esinesid vanused kategooriliselt, nii et andmestike vahel andmete ühtlustamiseks kategoriseerisime vanused ka teistes andmestikes samamoodi nagu RAHA andmestikus. RAHA ja ka edaspidi meie uuringus jagunes vanus 14 kategooriasse, kus 1 - <18, 2 – 18-19, 3 – 20-24, 4 – 25-29, 5 – 30-34, 6 – 35-49, 7 – 40-44, 8 – 45-49, 9 – 50-54, 10 – 55-59, 11 – 60-64, 12 – 65-69, 13 – 70-74 ja 14 - >75. Tabelis 2 on näha kärpimise järgselt alles jäänud andmete demograafilisi tunnuseid ja protsentuaalseid osakaale.

Tabel 2*Sotsiodemograafiline statistika*

Valim	Geenivaramu		Raha		Salk		RVTU	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Sugu								
Naine	54470	70.5	515	51.1	547	54.7	2780	61.8
Mees	22774	29.5	490	48.9	453	45.3	1715	38.2
Haridus								
Kõrgharidusega	44814	58	321	31.8	337	33.7	1422	31.7
Kõrghariduseta	32430	42	687	68.2	663	66.3	3065	68.3
Vanus								
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
	47.50	14.60	45.94	14.75	48.57	17.82	55.72	18.34
Alla 18	0	0	0	0	0	0	0	0
18-19	3413	4.4	11	1.1	24	2.4	132	2.9
20-24	5265	6.8	44	4.4	59	5.9	143	3.2
25-29	8588	11.1	115	11.4	64	6.4	179	4.0
30-34	8606	11.1	59	5.9	98	9.8	256	5.7
35-39	8637	11.2	153	15.2	79	7.9	250	5.6
40-44	8946	11.6	88	8.8	109	10.9	286	6.3
45-49	8960	11.6	106	10.5	85	8.5	360	8.0
50-54	7698	10.0	97	9.6	87	8.7	405	9.0
55-59	6407	8.3	84	8.4	78	7.8	450	10.0
60-64	4929	6.4	103	10.2	73	7.3	463	10.3
65-69	3407	4.4	88	8.8	67	6.7	422	9.4
70-74	1602	2.1	57	5.7	62	6.2	355	7.9
75 ja rohkem	786	1.0	0	0.0	115	11.5	794	17.7

Keskmete võrdlus

Student t-test

Meie viisime läbi 100 erinevat võrdlust Student *t*-testiga, millest kuus olid RAHA ja SALK andmestiku vahel, 22 EstBB ja RAHA andmestiku vahel, 28 EstBB ja SALK andmestiku vahel, 19 RVTU ja Raha andmestiku vahel, 6 RVTU ja SALK andmestiku vahel ning 19 EstBB ja RVTU valimi vahel. Väited olid normaaljaotuslikud ning vastasid *t*-testi eeldustele.

Student *t*-testiga leidsime, et peaaegu kõik väited olid oluliselt erinevad kõikide andmestike vahel. SALK ja RAHA valimid erinesid oluliselt nelja ühise väite puhul kuuest, mille tõttu ei olnud võimalik luua ühist esinduslikku andmestikku. EstBB ja esinduslike valimite vahel esines 28 väite keskmise olulist erinevust 44st (EstBB ja RAHA puhul 13/22 ja EstBB ja SALK puhul 20/28). Kõik EstBB ja esinduslike valimite ühised ekstraversust hindavad väited olid oluliselt erinevate keskmistega, kuid teisi isiksuse aspekte mõõtvate väidete puhul esines erinevuste kogus varieeruvalt. Väikseim suhe keskmiste oluliste erinevuste ja väidete arvu vahel oli EstBB ja RAHA valimi ühiste meelekindlust puudutavat väidete puhul, kus väited olid oluliselt erinevad kahel väitel kuuest. Suurim suhe keskmiste oluliste erinevuste vahel, peale ekstraversust hindavate väidete oli EstBB ja SALK valimi ühiste neurootilisust hindavate väidete puhul, kus olulisi erinevusi esines viiel juhul kuuest.

RVTU andmestiku ja esinduslike andmestike vahel esines 22 ühise väite puhul 17 olulist erinevust keskmistes (RVTU ja RAHA puhul 17/22 ja RVTU ja SALK puhul 5/6). Kõik RVTU esinduslike valimite vahelised ühised väited olid olulise keskmise erinevuse avatust kogemusele ja meelekindlust mõõtvate väidete puhul. RAHA ja SALK erinesid varieeruv määral ülejäänud väidete võrdluse osas RVTU-ga (korrelatsioonid ja korrelatsioonide erinevusi saab lähemalt vaadata Tabelis 3). Ekstraversust hindavate väidete puhul oli SALK ja RVTU vahel kaks ühtivat väidet, mis olid mõlemad oluliselt erinevad, kuid vastupidiselt EstBB ja RAHA võrdlusele RVTU ja RAHA andmestike vahel ühtivad 4 ekstraversuse väidet ei olnud statistiliselt oluliselt erinevad.

Mitte-esinduslike valimite, RVTU ja EstBB, andmestike vahel esines 16 erinevust 19 ühise väite puhul. Kõik neurootilisust ja avatust kogemusele mõõtvad väited olid oluliselt

erinevate keskmistega, abivalmiduse väidete puhul kaks kolmest, meelekindluse ja ekstravertsuse puhul kolm neljast väitest.

Keskmiised olid andmestike vahel suhteliselt erinevad. Arvestades, et väited olid kuue pallilisel Likert skaalal esines mitmeid väiteid, kus andmestike vahel olid keskmiste erinevused üle 0.25 ehk veerand palli, mille pidasime tähenduslikuks erinevuseks kahe vastamisstiili vahel. Eriti suurte erinevustena esines esinduslike vs mitte-esinduslike andmestike võrreldes EstBB ja SALK vahel erinevused: E26 $M_1-M_2 = 0.67$, E08 (0.75), O28 (0.78) ja RVTU ja SALK E08 (0.77). Samas andmestiku kategoorias olevate andmestike vahel olid eriti suured erinevused RAHA ja SALK andmestike vahel E08 (0.872), O28 (0.635) ning EstBB ja RVTU vahel 2 väidet N14(0.52) ja N36(0.53). Joonisel 1 on visuaalselt kujutatud andmestike ühiste väidete keskmisi üksteise kõrval koos vastavalt märgitud oluliste erinevustega.

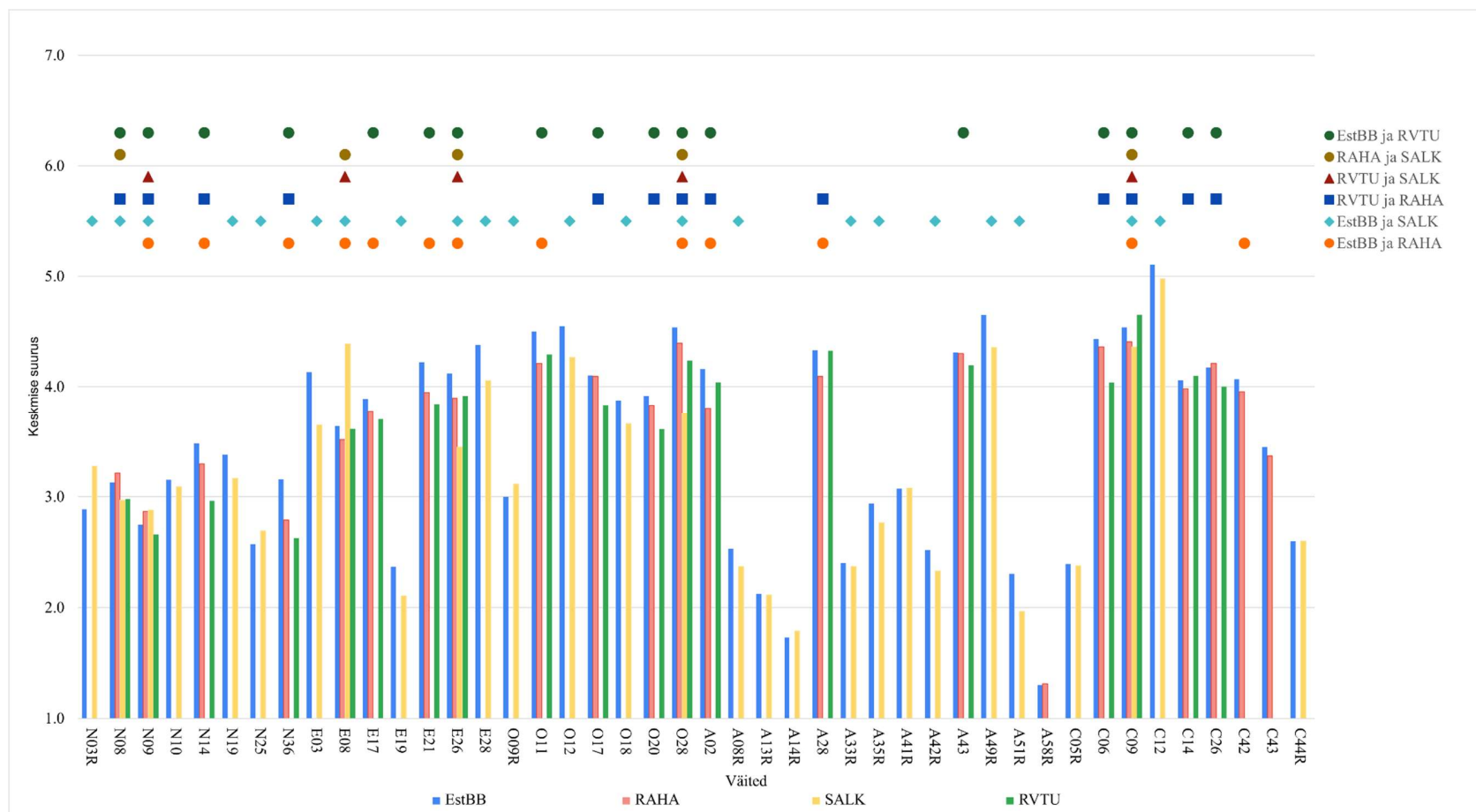
Tabel 3

Student t-testi kirjeldav tabel

Väide	EstBB		RAHA		SALK		RVTU		EstBB/ RAHA	EstBB/ SALK	RVTU/ RAHA	RVTU/ SALK	RAHA/ SALK	EstBB/ RVTU
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i> (78247)	<i>t</i> (78242)	<i>t</i> (5498)	<i>t</i> (5493)	<i>t</i> (2003)	<i>t</i> (81737)
N														
N03R	2.9	1.2			3.3	1.3				9.6*				
N08	3.1	1.2	3.2	1.2	3.0	1.2	3.0	1.2	2.2	4.1*	5.6*	0.2	4.5*	8*
N09	2.7	1.3	2.9	1.2	2.9	1.2	2.7	1.3	3.2*	3.3*	5*	5.1*	0.2	4.5*
N10	3.2	1.2			3.1	1.2				1.8				
N14	3.5	1.3	3.3	1.3			3.0	1.4	4.4*		7.3*			24.7*
N19	3.4	1.3			3.2	1.3				5.1*				
N25	2.6	1.2			2.7	1.2				3.2*				
N36	3.2	1.5	2.8	1.4			2.6	1.5	8.4*		3.3*			23.4*
E														
E03	4.1	1.2			3.7	1.2				12.4*				
E08	3.6	1.3	3.5	1.2	4.4	1.1	3.6	1.3	3.2*	21.6*	2.3	19.7*	17*	1.3
E17	3.9	1.1	3.8	1.1			3.7	1.2	3.4*		1.9			10.2*
E19	2.4	1.2			2.1	1.2				7.1*				
E21	4.2	1.1	3.9	1.1			3.8	1.2	7.6*		2.7			20.4*
E26	4.1	1.0	3.9	1.0	3.4	1.2	3.9	1.1	6.8*	17*	0.5	10.9*	8.8*	11.7*
E28	4.4	1.1			4.1	1.1				9.1*				
O														
O09R	3.0	1.3			3.1	1.3				2.8				
O11	4.5	1.3	4.2	1.4			4.3	1.5	6.6*		1.7			9.1*
O12	4.5	1.0			4.3	1.0				8.4*				
O17	4.1	1.1	4.1	1.1			3.8	1.3	0.3		6.9*			14.2*
O18	3.9	1.3			3.7	1.3				5.2*				
O20	3.9	1.2	3.8	1.2			3.6	1.3	2.3		5.4*			15.8*

Väide	EstBB		RAHA		SALK		RVTU		EstBB/ RAHA	EstBB/ SALK	RVTU/ RAHA	RVTU/ SALK	RAHA/ SALK	EstBB/ RVTU
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i> (78247)	<i>t</i> (78242)	<i>t</i> (5498)	<i>t</i> (5493)	<i>t</i> (2003)	<i>t</i> (81737)
O28	4.5	1.0	4.4	1.1	3.8	1.0	4.2	1.1	4.3*	23.7*	4.1*	13*	13.6*	17.2*
A														
A02	4.2	1.0	3.8	1.1			4.0	1.1	10.6*		6.4*			7.4*
A08R	2.5	1.2			2.4	1.2				4.2*				
A13R	2.1	1.1			2.1	1.1				0.2				
A14R	1.7	1.0			1.8	1.0				1.7				
A28	4.3	1.0	4.1	1.1			4.3	1.1	7.1*		6.3*			0.3
A33R	2.4	1.2			2.4	1.2				0.8*				
A35R	2.9	1.2			2.8	1.2				4.6*				
A41R	3.1	1.2			3.1	1.1				0.2				
A42R	2.5	1.0			2.3	1.0				6*				
A43	4.3	1.1	4.3	1.1			4.2	1.2	0.2		2.8			6.4*
A49R	4.7	1.0			4.4	1.1				8.3*				
A51R	2.3	1.3			2.0	1.2				8.7*				
A58R	1.3	0.7	1.3	0.7					0.6					
C														
C05R	2.4	1.1			2.4	1.2				0.3				
C06	4.4	1.1	4.4	1.1			4.0	1.3	2.1		8.3*			19.5*
C09	4.5	1.0	4.4	1.0	4.4	1.1	4.7	1.0	4.1*	5*	7*	7.7*	1.0	7.7*
C12	5.1	0.8			5.0	0.8				4.7*				
C14	4.1	1.2	4.0	1.1			4.1	1.2	2.1		3*			2.3
C26	4.2	1.2	4.2	1.1			4.0	1.4	1.2		5.3*			8.3*
C42	4.1	1.3	4.0	1.2					3*					
C43	3.5	1.4	3.4	1.3					2.1					
C44R	2.6	1.1			2.6	1.1				0				

Märkus: * $p < 0.005$.



Joonis 1

Kõikide andmestike väidete keskmised ja Student t-testi tuvastatud olulised erinevused andmestike vahel

Märkus: Täpid tähistavad vastavate andmestike vahelisi statistiliselt olulisi erinevusi. Üle 0.25 pallist keskmiste erinevust esines andmestike vahel väidetes N03R, N14, N36, E03, E08, E19, E21, E26, E28, O11, O12, O17, O20, O28, A02, A49R, A51R, C06 ja C09.

Korrelatsioonid

Sugu. Leidsime, et sootunnuse korrelatsioonid erinesid oluliselt kaheksal võrdlusel 100st. Esinduslike ja mitte-esinduslike valimite vahel esines 5 olulist korrelatsiooni erinevust, millest väide A58R oli isegi eri suundadega. EstBB ja RAHA andmestike vahel esines oluline korrelatsiooni erinevus 2 väite puhul, N36 ($r_{EstBB} = 0.37$, $r_{RAHA} = 0.44$, $z = 2.61$, $p < .005$) ja erisuundadega väide A58R ($r_{EstBB} = -0.12$, $r_{RAHA} = 0.11$, $z = 7.15$, $p < .005$). EstBB ja SALK andmestike vahel esines kolmes väites oluline erinevus, A14R ($r_{EstBB} = -0.21$, $r_{SALK} = -0.11$, $z = 3.35$, $p < .005$), A33R ($r_{EstBB} = -0.17$, $r_{SALK} = -0.09$, $z = 2.65$, $p < .005$) ja A35R ($r_{EstBB} = -0.18$, $r_{SALK} = -0.1$, $z = 2.72$, $p < .005$). EstBB ja RVTU andmestike vahel esines erinevus kolmes väites N09 ($r_{EstBB} = 0.05$, $r_{RVTU} = 0.12$, $z = 4.46$, $p < .005$), E26 ($r_{EstBB} = 0.1$, $r_{RVTU} = -0.04$, $z = 3.26$, $p < .005$) ja C09 ($r_{EstBB} = 0.07$, $r_{RVTU} = 0.12$, $z = 3.76$, $p < .005$). Jälle saab täheldada erisuunalist korrelatsiooni väite E26 puhul. RAHA ja SALK, RVTU ja RAHA ning RVTU ja SALK võrdluste puhul ei esinenud korrelatsioonide olulist erinevust. Otsustasime sootunnuste korrelatsioonide erinevusi raporteerida ainult tekstisisese korrelatsioonide erinevuste vähesuse tõttu.

Mitte-esinduslike ja esinduslike valimite võrdlusel, kui esines oluline korrelatsiooni erinevus, ühtis see sama andmestike keskmiste erinevuse võrdlusega kahel juhul viiest. EstBB ja RAHA andmestike vahel väide N36, EstBB ja SALK andmestike vahel A35R. Ülejäänud kolm väidet olid korrelatsiooni erinevusega, kuid keskmised ei olnud oluliselt erinevad. Kõikide EstBB ja RVTU andmestike korrelatsioonide erinevustega ühtis ka keskmiste oluline erinevus.

Haridus. Hariduse korrelatsioonid erinesid oluliselt 15 võrdluse puhul. Esinduslike ja mitte-esinduslike andmestike vahel esines neist 5. EstBB ja RAHA andmestike vahel erinesid väited C42 ja C43 ning EstBB ja SALK andmestike vahel N10 ja N19 (Tabelis 4 saab näha hariduse korrelatsioone ja erinevusi täpsemalt). RVTU ja RAHA andmestike vahel esines oluline erinevus vaid väites C06 ning RVTU ja SALK andmestike vahel ei esinenud olulist erinevust. Taaskord EstBB ja RAHA andmestike vahel esines erineva suunaga korrelatsioon, seekord väite C42 puhul. RAHA ja SALK andmestikud ei erinenud oluliselt, kuid EstBB ja RVTU korrelatsioonid erinesid 10 väite puhul. Erineva suunaga korrelatsioon esines EstBB ja RVTU vahel väitel N14.

Tabel 4

Hariduse korrelatsioonid ja korrelatsioonide erinevusi kirjeldav tabel

Tunnus	EstBB <i>r</i> (77244)	RAHA <i>r</i> (1005)	SALK <i>r</i> (1000)	RVTU <i>r</i> (4495)	1	2	3	4	5	6
					<i>z</i>					
N										
N03R	0.01		0.07			1.85				
N08	-0.05*	0	-0.07	-0.04*	1.48	0.76	1.07	0.97	1.59	0.64
N09	-0.07*	-0.02	-0.12*	-0.08*	1.45	1.46	1.59	1.06	2.07	0.61
N10	-0.05*		-0.14*			2.79*				
N14	0.03*	0.02		-0.02	0.06		1.27			3.03*
N19	-0.01		-0.11*			3.21*				
N25	-0.14*		-0.15*			0.61				
N36	-0.03*	-0.02		-0.05*	0.32		0.81			1.19
E										
E03	0.04*		0.04			0.03				
E08	0	-0.02	0.03	0.04*	0.54	0.91	1.76	0.44	1.03	2.9*
E17	0.05*	0.04		0.11*	0.35		2.21			4.32*
E19	-0.02*		0			0.53				
E21	0	0		0.06*	0		1.74			3.96*
E26	0.04*	-0.02	0.03	0.07*	1.92	0.35	2.49	1.05	1.12	1.68
E28	0.02*		0.06			1.1				
O										
O09R	-0.14*		-0.13*			0.35				
O11	0.17*	0.2*		0.19*	0.85		0.11			1.5
O12	0.12*		0.08			1.17				
O17	0.02*	0.03		0.08*	0.54		1.32			4.11*
O18	0.09*		0.06			0.95				
O20	0.13*	0.13*		0.17*	0.13		1.12			2.8*
O28	0.08*	0.12*	0.1*	0.13*	0.99	0.54	0.47	0.87	0.32	3.1*
A										
A02	0.1*	0.1*		0.1*	0.1		0.08			0.38
A08R	0.01*		0			0.41				
A13R	0.01		0.05			1.38				
A14R	-0.07*		-0.02			1.7				
A28	0.02*	0.05		0.09*	1.07		1.01			4.5*
A33R	-0.03*		-0.03			0.06				
A35R	0.11*		0.1*			0.19				
A41R	0.04*		0.02			0.44				

Tunnus	EstBB <i>r</i> (77244)	RAHA <i>r</i> (1005)	SALK <i>r</i> (1000)	RVTU <i>r</i> (4495)	1	2	3	4	5	6
A42R	-0.02*		0.03			1.38				
A43	0.04*	0.05		0.1*	0.38		1.48			4.14*
A49R	0.1*		0.06			1.2				
A51R	-0.04*		-0.05			0.25				
A58R	-0.05*	-0.04			0.54					
C										
C05R	-0.07*		0.01			2.32				
C06	0.08*	0		0.13*	2.37		3.57*			3.21*
C09	0	0.07	-0.04	0.03	2.11	1.19	1.21	1.8	2.35	1.62
C12	0.07*		0.03			1.23				
C14	-0.04*	-0.01		0	1.04		0.17			2.54
C26	0.03*	0.02		0.04*	0.44		0.79			0.89
C42	0.08*	-0.07			4.64*					
C43	-0.07*	0.06			4.23*					
C44R	-0.07*		-0.04			0.94				

Märkus: * $p < 0.005$. 1 – EstBB ja RAHA, 2 – EstBB ja SALK, 3 – RVTU ja RAHA, 4 – RVTU ja SALK, 5 - RAHA ja SALK, 6 – EstBB ja RVTU.

Mitte-esinduslike ja esinduslike valimite võrdlusel, kui esines oluline korrelatsiooni erinevus, ühtis see sama andmestike keskmiste erinevusega kolmel juhul viiest. EstBB ja RAHA andmestike vahel väide C42, EstBB ja SALK andmestike vahel N19 ja RVTU ja RAHA vahel C06. Ülejäänud kolm väidet olid korrelatsiooni erinevusega, kuid keskmised ei olnud oluliselt erinevad. Kõik EstBB ja RVTU andmestike korrelatsiooni erinevused ühtisid keskmiste olulise erinevusega välja arvatud kahel juhul 10st, väited A28 ja E08 olid korrelatsiooni erinevusega, kuid ilma keskmiste olulise erinevusega.

Vanus. Vanuse korrelatsioonid erinesid oluliselt 20 võrdluse puhul. Esinduslike ja mitte-esinduslike andmestike vahel esines neist 8. EstBB ja RAHA andmestike vahel erinesid oluliselt väited C26 ja A58R ning EstBB ja SALK andmestike vahel väited O18 ja A13R. RVTU ja RAHA andmestike vahel erinesid väited E17, O20, C06 ja C26 ning RVTU ja SALK andmestike vahel ei esinenud olulisi erinevusi (Korrelatsioone ja erinevusi on võimalik näha Tabelis 5). Mitte-esinduslike ja esinduslike andmestike vahel erineva suunaga korrelatsioonid esinesid väite A58R ja E17 puhul. RAHA ja SALK andmestike vahel ei esinenud olulisi erinevusi, kuid EstBB

ja RVTU vahel esines 10 oluliselt erinevat korrelatsiooni. Mitte-esinduslike andmestike vahel erinevate suundadega korrelatsioone esines väidete E17 ja C26 puhul.

Tabel 5

Vanuse korrelatsioone ja korrelatsioonide erinevusi kirjeldav tabel

Tunnus	EstBB	RAHA	SALK	RVTU	1	2	3	4	5	6
	<i>r</i> (77244)	<i>r</i> (1005)	<i>r</i> (1000)	<i>r</i> (4495)	<i>z</i>					
N										
N03R	0.13*		0.06			2.12				
N08	-0.07*	-0.05	-0.11*	-0.12*	0.92	1.11	2.21	0.36	1.44	3.13*
N09	-0.16*	-0.12*	-0.15*	-0.1*	1.19	0.45	0.58	1.25	0.52	3.78*
N10	-0.21*		-0.18*			1.01				
N14	-0.22*	-0.16*		-0.17*	1.86		0.34			3.07*
N19	-0.19*		-0.16*			1.16				
N25	-0.04*		-0.08			1.13				
N36	-0.08*	-0.06		-0.02	0.79		0.99			3.89*
E										
E03	-0.04*		-0.05			0.19				
E08	0	0.06	-0.01	0.01	1.92	0.53	1.55	0.69	1.75	0.46
E17	0.04*	0.05		-0.08*	0.32		3.71*			7.77*
E19	-0.16*		-0.18*			0.58				
E21	-0.17*	-0.19*		-0.25*	0.59		1.76			5.22*
E26	-0.15*	-0.15*	-0.2*	-0.15*	0	1.68	0.06	1.48	1.2	0.13
E28	-0.07*		-0.07			0.13				
O										
O09R	0.1*		0.12*			0.51				
O11	0.15*	0.17*		0.17*	0.52		0.05			1.18
O12	-0.1*		-0.11*			0.19				
O17	0.02*	-0.03		-0.06*	1.58		0.84			5.16*
O18	-0.03*		-0.15*			3.93*				
O20	-0.08*	-0.01		-0.1*	2.21		2.6*			1.34
O28	-0.01	-0.02	0.01	0.03	0.25	0.5	1.45	0.76	0.54	2.78*
A										
A02	0.05*	0.08		0.11*	0.73		1.08			3.95*
A08R	-0.23*		-0.3*			2.37				
A13R	-0.09*		-0.19*			3.1*				
A14R	-0.12*		-0.12*			0.03				
A28	0.07*	0.1*		0.04	0.95		1.82			2.17
A33R	-0.05*		-0.06			0.16				

Tunnus	EstBB	RAHA	SALK	RVTU	1	2	3	4	5	6
	<i>r</i> (77244)	<i>r</i> (1005)	<i>r</i> (1000)	<i>r</i> (4495)	<i>z</i>					
A35R	-0.15*		-0.15*			0.16				
A41R	-0.13*		-0.07			1.93				
A42R	-0.09*		-0.11*			0.76				
A43	0.08*	0.05		0.04*	0.98		0.28			2.66*
A49R	-0.22*		-0.17*			1.47				
A51R	-0.23*		-0.21*			0.49				
A58R	-0.05*	0.06			3.31*					
C										
C05R	-0.02*		-0.07			1.54				
C06	-0.13*	-0.12*		-0.31*	0.45		5.73*			12.1*
C09	0.08*	0.01	0.02	0.09*	2.18	1.6	2.5	1.97	0.4	1.17
C12	0.07*		0			2.01				
C14	0.16*	0.13*		0.13*	1.19		0.1			2.23
C26	0.05*	0.15*		-0.01	2.96*		4.61*			4.36*
C42	0.08*	0.04			1.3					
C43	-0.17*	-0.19*			0.39					
C44R	-0.1*		-0.06			1.39				

Märkus: * $p < 0.005$. 1 – EstBB ja RAHA, 2 – EstBB ja SALK, 3 – RVTU ja RAHA, 4 – RVTU ja SALK, 5 - RAHA ja SALK, 6 – EstBB ja RVTU.

Mitte-esinduslike ja esinduslike valimite võrdlusel, kui esines oluline korrelatsiooni erinevus ühtis see samade andmestike keskmiste erinevusega neljal juhul kaheksast. EstBB ja SALK andmestike vahel O18 ja RVTU ja RAHA vahel O20, C06 ja C26. Ülejäänud neli väidet olid korrelatsiooni erinevusega, kuid keskmised ei olnud oluliselt erinevad. EstBB ja RVTU andmestike korrelatsiooni erinevused ühtisid keskmiste olulise erinevusega.

Arutelu

Isiksuse-uuringute mitteesinduslikkusest räägitakse palju, kuid seda uuritakse üsna harva. Käesolev uuringus haaras nelja erinevat valimit – kahe esinduslikku ja kahe mitte-esinduslikku, et esinduslikkuse probleemi hinnata. Leidsime, et isegi esinduslikud valimid erinesid omavahel isiksuseväidete osas, rääkimata siis erinevustest mitte-esinduslike ja esinduslike valimite vahel. Uurisime ka isiksuse ja demograafia seoseid erinevatel valimitel. Ehkki need erinesid, ei erinenud nad enamasti samadel väidetest kus keskmised erinesid. Seega on raske ütelda, kas keskmiste erinevus otseselt seoseid mõjutab.

Esimese uurimisküsimuse vastus

Esimesele uurimisküsimusele selget vastust ei olnud võimalik leida. Kuigi esinduslike ja mitte-esinduslike väidete keskmiste vahel esines suures koguses olulisi erinevusi, esines ka esinduslike valimite enda vahel olulisi erinevusi keskmistes. Osaluskalde tõestusviis näitab, et osaluskalle võis esineda EstBB puhul 28 ja RVTU puhul 17 väite puhul, kuid selle seab kahtluse alla esinduslike valimite omavaheline erinevus. Kuna kõik valimid olid omavahel erinevad on raske seada osaluskallet põhiliseks erinevuste põhjustajaks. Osaluskalde tuvastamise loogikat kasutades, et keskmiste erinevus peegeldab osaluskalde esinemist, peaks üks esinduslikest valimitest olema ka osaluskaldega ning antud uuringu raames ei ole võimalik tuvastada, milline esinduslik valim peaks olema tegelik esinduslikkust kehastav võrdluspunkt.

Teine võimalik tõlgendus on, et vaatamata esinduslike valimite eesmärgile olla tasakaalus soo, vanuse, hariduse ja keele osas, ei ole nad ikkagi tasakaalus isiksuse osas. Võimalik, et erinevate uuringufirmade meetodid apelleerivad erineva isiksusega inimestele ja nii on paneelid erinevate hoiakutega (Jüristo, 2021; Mölder, 2021).

Teise uurimisküsimuse vastus

Tulemused andsid teise uurimisküsimuse osas selgema vastuse. Arvestades keskmiste erinevuste paljusust oli imestusväärset vähe erinevusi korrelatsioonides, eriti esinduslike ja mitte-esinduslike valimite vahel. Kuigi sarnaselt Van Alten jt. (2022) leidsime, et esines erisuunalisi korrelatsioone andmestike vahel, oli tegemist üksikute väidetege vähestest oluliselt erineva demograafilise korrelatsiooniga väidetest. Lisaks korrelatsioonide erinevuse vähesusele, olid vaid pooled või vähemad esinevatest erinevustest, sõltuvalt demograafilisest tunnusest, kattuvad keskmiste erinevusega. Juhul kui meie mitte-esinduslikes valimites oli osaluskalle siis isiksuste erinevused laiaulatuslikku mõju ei tundunud rakendavat uuringute tulemustele. See tulemus läheb kokku Sogaard jt. (2004) tööga, kus osaluskalle ei mõjutanud oluliselt korrelatsiooni, kuid vastu Schoeler jt. (2023) uuringule, kus korrelatsioonid olid tugevalt mõjutatud osaluskaldest.

Huvitaval kombel kahe esindusliku valimi vahel ei esinenud olulisi korrelatsiooni erinevusi, kuigi neil esines pooltes ühistes väidetes oluliselt erinevad keskmised. Kõige rohkem korrelatsiooni erinevusi esines just kahe mitte-esindusliku uuringu vahel. See võib-olla tingitud

sellest, et mõlemad valimid olid vabatahtlikud ja erineva suurusega. Vastupidiselt kahele esinduslikule valimile, on oodatavam, et mitte-esinduslikud valimid erinevad andmete keskmiste ja korrelatsioonide osas. Samas olid mitte-esinduslikud valimid ka oluliselt suuremad, seega tulid väiksemad erinevused statistiliselt oluliselt välja. Lisaks koguti need üsna ühtemoodi - enamjaoolt veebi kutsetega.

Puudused ja kokkuvõte

Meie tööd piiras oluliselt ühtse võrdluspunkti puudumine. Esinduslikud valimid oleksid pidanud algselt moodustama suurema esinduslike väitetulemuste kogumi, kuid keskmised ei klappinud piisavalt, et seda lubada. Uurimaks isiksusjoonte mõju osaluskaldele, võiks tulevastes uuringutes kindlasti kasutada mitte-esinduslike uuringutega ulatuslikult kattuvate väidetega esinduslikku valimit. Lisaks tuleks kasutada statistilisi kalde korrigeerimise meetodeid, nagu näiteks kaalumine, et vaadata, kas kalde korrigeerimine korrigeerib ka korrelatsioone ühtlasemas suunas.

Kokkuvõttes erinesid andmestikud laialdaselt väidete keskmiste osas, näidates isiksusjoonte tugevuste suurt erinevust valimite vahel. Kindlaid osaluskaldega väiteid ei ole võimalik aga välja tuua esinduslike valimite erinevuste tõttu. Sõltumata raskusest omistada osaluskallet spetsiaalsetele väidetele, leidsime, et isiksusjoonte keskmiste oluline erinevus ei ennustanud demograafiliste korrelatsioonide olulist erinevust mitte-esinduslike ja esinduslike valimite vahel.

Kasutatud allikad

- Arumäe, K., Vaht, M., Realo, A., Ausmees, L., Allik, J., Henry, S., Metspalu, A., Esko, T., Mõttus, R. & Vainik, U. (2025, 3. veebruar). Cohort Profiles: Personality Measurements at the Estonian Biobank of the Estonian Genome Center, University of Tartu. https://doi.org/10.31234/osf.io/2aey6_v2
- Bosnjak, M., Haas, I., Galesic, M., Kaczmirek, L., Bandilla, W., & Couper, M. P. (2013). Sample Composition Discrepancies in Different Stages of a Probability-based Online Panel. *Field Methods*, 25(4), 339-360. <https://doi.org/10.1177/1525822X12472951>
- Comparing Correlation Coefficients*. (2025, 16. jaanuar). Statistics Solutions. <https://www.statisticssolutions.com/comparing-correlation-coefficients/>
- Jüristo, T. (25 August, 2021). Valed, neetud valed ja statistika – Salk. *Salk*. <https://salk.ee/artiklid/valed-neetud-valed-ja-statistika/>
- Marcus, B., & Schütz, A. (2005). Who are the people reluctant to participate in research? Personality correlates of four different types of nonresponse as inferred from self- and observer ratings. *Journal of personality*, 73(4), 959–984. <https://doi.org/10.1111/j.1467-6494.2005.00335.x>
- Milani, L., Alver, M., Laur, S., Reisberg, S., Haller, T., Aasmets, O., Abner, E., Alavere, H., Allik, A., Annilo, T., Fischer, K., Hofmeister, R., Hudjashov, G., Jöeloo, M., Kals, M., Karo-Astover, L., Kasela, S., Kolde, A., Krebs, K., . . . Metspalu, A. (2025). The Estonian Biobank’s journey from biobanking to personalized medicine. *Nature Communications*, 16(1). <https://doi.org/10.1038/s41467-025-58465-3>
- Murakas, R. U. (2025). A Multidimensional View of Financial Well-being and the Links to Personality: An Explorative Research. Tartu Ülikool. <https://dspace.ut.ee/items/662fcc45-a155-4d4f-bb0b-fd92559ed6a1>
- Mölder, M. (2021, 8 detsembril). *Erinevused erakondade toetusküsitlustes - Eesti Poliitika*. Eesti Poliitika. <https://eestipoliitika.ee/blogi/erinevused-erakondade-toetuskusitlustes/>

- Pagan, J.L., Eaton, N.R., Turkheimer, E., & Oltmanns, T.F. (2006). Peer-reported personality problems of research nonparticipants: Are our samples biased? *Personality and Individual Differences*, *41*, 1131-1142. <https://doi.org/10.1016/j.paid.2006.04.017>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological methods*, *17*(3), 354–373. <https://doi.org/10.1037/a0029315>
- Robitzsch, A. (2020). Why Ordinal Variables Can (Almost) Always Be Treated as Continuous Variables: Clarifying Assumptions of Robust Continuous and Ordinal Factor Analysis Estimation Methods. *Frontiers in Education*, *5*.
<https://doi.org/10.3389/feduc.2020.589965>
- Schoeler, T., Speed, D., Porcu, E., Pirastu, N., Pingault, J. & Kutalik, Z. (2023). Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nat Hum Behav* *7*, 1216–1227. <https://doi.org/10.1038/s41562-023-01579-9>
- Søgaard, A.J., Selmer, R., Bjertness, E. & Thelle, D. (2004). The Oslo Health Study: The impact of self-selection in a large, population-based survey. *Int J Equity Health* *3*, 3.
<https://doi.org/10.1186/1475-9276-3-3>
- Van Alten, S., Domingue, B. W., Galama, T., & Marees, A. T. (2022). Reweighting the UK Biobank to reflect its underlying sampling population substantially reduces pervasive selection bias due to volunteering. *medRxiv (Cold Spring Harbor Laboratory)*. <https://doi.org/10.1101/2022.05.16.22275048>

Käesolevaga kinnitan, et olen korrektselt viidanud kõigile oma töös kasutatud teiste autorite poolt loodud kirjalikele töödele, lausetele, mõtetele, ideedele või andmetele.

Olen nõus oma töö avaldamisega Tartu Ülikooli digitaalarhiivis DSpace.

Raul Jack Barrow