

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Hanna Kangur
**Kodade virvendusarütmia ja laperduse riski
prognoosimine TÜ Eesti Geenivaramu näitel**

Matemaatilise statistika eriala
Bakalaureusetöö (9 EAP)

Juhendajad: Prof. Krista Fischer, PhD
Tuuli Puusepp, MSc

TARTU 2026

Kodade virvendusarütmia ja laperduse riski prognoosimine TÜ Eesti Geenivaramu näitel

Bakalaureusetöö

Hanna Kangur

Lühikokkuvõte

Bakalaureusetöö eesmärk on selgitada välja kodade virvendusarütmia ja laperduse mittegeneetilised riskitegurid ning uurida geneetilise riski mõju kodade virvendusarütmia ja laperduse avaldumisele. Riskitegurite tuvastamiseks ja geneetilise riski mõju uurimiseks rakendatakse Tartu Ülikooli Eesti Geenivaramu andmetel elukestusanalüüsi meetodeid nagu Coxi võrdeliste riskide mudel ja Aalen-Johanseni hinnang. Lisaks võrreldakse Harrelli C-indeksi alusel mittegeneetiliste tegurite ja geneetilise riskiskoori suhtelist olulisust kodade virvendusarütmia ja laperduse tekkeriski prognoosimisel. Analüüsi tulemusena selgus, et kodade virvendusarütmia ja laperduse tekkerisk on oluliselt seotud mitmete mittegeneetiliste riskiteguritega, aga ka geneetilise riskiskooriga, kusjuures geneetiline eelsoodumus osutus kodade virvendusarütmia ja laperduse tugevaimaks üksikuks ennustajaks.

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

Märksõnad: elukestusanalüüs, prognostilised mudelid, polügeenne riskiskoor, kodade virvendusarütmia ja laperdus

**Risk prediction of atrial fibrillation and flutter using data from the
Estonian Genome Centre at the University of Tartu**

Bachelor's thesis

Hanna Kangur

Abstract

The aim of the bachelor's thesis is to identify non-genetic risk factors for atrial fibrillation and flutter and to examine the impact of genetic risk on the onset of these arrhythmias. Survival analysis methods, including the Cox proportional hazards model and the Aalen-Johansen estimator, are applied to data from the Estonian Genome Centre at the University of Tartu. In addition, the relative importance of both non-genetic factors and the genetic risk score in predicting the risk of developing atrial fibrillation and flutter is compared using Harrell's C-index. The results show that several non-genetic risk factors as well as the genetic risk score are significantly associated with the risk of developing atrial fibrillation and flutter, with genetic predisposition being the strongest single predictor of these arrhythmias.

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

Keywords: survival analysis, prognostic models, polygenic risk score, atrial fibrillation and flutter

Sisukord

Sissejuhatus	6
1 Kodade virvendusarütmia ja laperdus	8
2 Polügeenne riskiskoor	10
3 Elukestusanalüüs	12
3.1 Tsenseeritus	12
3.1.1 Informatiivne ja mitteinformatiivne tsenseerimine	13
3.2 Ajaskaala valik	14
3.3 Üleelamis- ja riskifunktsioon	14
3.4 Kaplan-Meieri hinnang	15
3.5 Konkureerivad riskid	17
3.5.1 Aalen-Johanseni hinnang	18
3.6 Coxi võrdeliste riskide mudel	18
3.7 Võrdeliste riskide eelduse kontrollimine	20
3.8 Kihitatud Coxi võrdeliste riskide mudel	21
3.9 Harrelli C-indeks	22
4 Logistiline regressioonimudel	24
5 Andmete analüüs	25
5.1 Ülevaade andmetest	25
5.2 PRSi valimine	26
5.3 Kirjeldav analüüs	27
5.4 Kodade virvendusarütmia ja laperduse riskitegurid	31
5.4.1 KVA risk PRSi protsentiilgruppides	32

5.4.2	Coxi võrdeliste riskide mudelid	35
5.4.3	Mudelite võrdeliste riskide eelduse kontrollimine	40
5.4.4	C-indeksite võrdlus	40
6	Arutelu	42
	Kokkuvõte	44
	Kasutatud allikad	46
	Lisad	50

Sissejuhatus

Bakalaureusetöö eesmärk on Tartu Ülikooli Eesti Geenivaramu andmete põhjal välja selgitada kodade virvendusarütmia ja laperduse tekkeriskiga seotud mittegeneetilised riskifaktorid ning hinnata geneetilise riskiskoori lisandväärtust prognoosimudelites.

Kodade virvendusarütmia ja laperdus kuuluvad levinumate südame rütmihäirete hulka ning kuigi need ei ole otseselt eluohtlikud, võivad need põhjustada tõsisid tüsistusi, millest olulisemad on insult ja südamepuudulikkus (Järving, Teeäär, Elmet, Laane, Arus, 2021). Kaasaegsed diagnostikavahendid võimaldavad südame rütmihäireid varakult tuvastada ja ravida, kuid nende kindlakstegemiseks vajalikke uurin-guid viiakse enamasti läbi vaid selge kahtluse või teadaolevate riskitegurite esinemise korral (Van Gelder *et al.*, 2024). Sellest tulenevalt võivad vanuse, elustiili ja kaasuva-te haiguste põhjal näiliselt madala riskiga inimesed jääda diagnoosita ning seetõttu ilma õigeaegsest sekkumisest, mis viitab täpsemate riskihindamise meetodite vaja-dusele.

Traditsioonilised prognoosimudelid keskenduvad valdavalt mittegeneetilistele riski-teguritele nagu vanus, suitsetamine, ülekaalulisus ja krooniliste haiguste esinemine (Alonso *et al.*, 2013). Viimaste aastate teadusuuringud on aga kinnitanud, et koda-de virvendusarütmia ja laperduse kujunemisel mängib olulist rolli ka geneetiline eel-soodumus (Roselli, Rienstra, Ellinor, 2020). Seetõttu võib geneetilise riskiskoori kaa-samine, kombineerituna klassikaliste riskifaktoritega, oluliselt parandada prognoosi-mudelite täpsust ja aidata tuvastada riskigruppe, kes jääksid muidu tähelepanuta (Segan *et al.*, 2025).

Töö jaguneb teoreetiliseks ja praktiliseks osaks. Teoreetiline osa koosneb neljast pea-tükist. Neist esimeses käsitletakse kodade virvendusarütmia ja laperduse olemust ning riskitegureid, teises tutvustatakse polügeenset riskiskoori. Kolmandas peatükis antakse ülevaade töös kasutatavatest elukestusanalüüsi meetoditest, sealhulgas nii klassikalistest meetoditest nagu Kaplan-Meieri hinnang ja Coxi võrdeliste riski-de mudel kui ka konkureerivaid riske arvesse võtvast Aalen-Johanseni hinnangust. Neljandas peatükis kirjeldatakse logistilise regressiooni mudelit, mida rakendatak-se elukestusanalüüsis kasutatava geneetilise riskiskoori valimiseks. Praktilises osas

antakse ülevaade Tartu Ülikooli Eesti Geenivaramu (edaspidi ka TÜ Eesti Geenivaramu või geenivaramu) andmetest ja nende põhjal tehtud analüüsi tulemustest.

1 Kodade virvendusarütmia ja laperdus

Süda koosneb neljast kambrist, millest ülemisi nimetatakse kodadeks ja alumisi vatsakesteks. Vere pumpamiseks kehas tõmbub südamelihaski siinussõlmest lähtuvate elektriimpulsside mõjul kokku. Normaalse südame töö puhul tõmbuvad regulaarselt kokku esmalt kojad ja seejärel vatsakesed. Südame rütmihäirete korral esineb aga kõrvalekalle südame normaalsest rütmist ehk siinusrütmist, mis võib väljenduda nii muutunud löögisageduse kui ka ebaregulaarse südame tegevusena. (Regionaalhaigla Südamekeskus, s. a.)

Kodade virvendusarütmia (KVA) korral liiguvad elektriimpulsid kodades ringi korrapäraselt, mistõttu ei tõmbu kojad enam koordineeritult kokku, vaid hakkavad kiiresti ja kaootiliselt võbelema ehk virvendama. Selle tulemusena aeglustub vere liikumine kodadest vatsakestesse. KVA võib avalduda kas lühiajaliste episoodide ehk paroksüsmidena, mille vahelisel ajal taastub siinusrütm, või püsiva rütmihäirena, mille korral on süda pidevalt rütmist väljas. Enamasti on KVA kroonilise iseloomuga ning aja jooksul muutuvad episoodid sagedasemaks ja püsivamaks. (Järving, Teeäär, Elmet, Laane, Arus, 2021)

Kodade laperduse korral löövad südame kojad tavapärasest kiiremini, kuid erinevalt virvendusarütmiaist siiski koordineeritult. Kuigi kodade laperdus on vähem levinud kui KVA, on nende sümptomid, riskitegurid ja võimalikud tüsistused suures osas sarnased. Lisaks esineb kodade laperdusega indiviididel märkimisväärne risk kodade virvendusarütmia kujunemiseks ka pärast ravi. (American Heart Association, 2025) Kuna kodade virvendusarütmia ja kodade laperdus on kliiniliselt suurel määral kattuvad ning on rahvusvahelise haiguste klassifikatsiooni (RHK) jaotistes koondatud ühise koodi (I48) alla, siis käsitletakse neid töös ühiselt ja edaspidi hõlmab KVA mainimine mõlemat rütmihäiret.

KVA on kõige sagedamini esinev südame rütmihäire. Selle üheks olulisemaks riskiteguriks on vanus – üle 50-aastastel inimestel suureneb KVA tekkerisk iga kümne aastaga ligikaudu kaks korda. (Ravijuhend, 2021) Lisaks mängib väga suurt rolli elustiil. KVA teket ja püsimist soodustavad nii suitsetamine, alkoholi liigtarvitamine, ülekaalulisus, tasakaalustamata toitumine kui ka stress. Samuti suurendavad KVA

riski mitmed kroonilised haigused, sealhulgas kõrgvererõhktõbi, südame isheemiatõbi, südamepuudulikkus, diabeet, uneapnoe ja kilpnäärme talitlushäired. (Järving, Teeäär, Elmet, Laane, Arus, 2021) Hiljutiste genoomika-alaste edusammude toel on arvukad uuringud kinnitanud, et kodade virvendusarütmial on ka oluline geneetiline komponent. Seetõttu on geneetiliste ja mittegeneetiliste tegurite integreerimine oluline täpsema ning terviklikuma riskihindamise saavutamiseks, eriti varajase avastamise ja personaalse ennetuse tagamiseks. (Andersen, Andreassen, Olesen, 2020)

Paljud inimesed elavad kodade virvendusarütmiauga pikka aega ja seisund ei ole otseselt eluohtlik. Küll aga võib see põhjustada tõsiseid tüsistusi, millest olulisemad on insult ja südamepuudulikkus. Kui kojad ei tõmbu regulaarselt kokku, mille tagajärjel ei pumbata neist piisavalt verd välja, võivad südames tekkida verehüübed ehk trombid, mis võivad veresoonte kaudu liikuda teistesse organitesse. Kui tromb jõuab ajuarteritesse, võib see põhjustada insulti. (Järving, Teeäär, Elmet, Laane, Arus, 2021) KVA diagnoosiga inimestel on insuldirisk ligikaudu viis korda suurem võrreldes nendega, kellel rütmihäiret ei esine (Migdady, Russman, Buletko, 2021).

2 Polügeenne riskiskoor

Geneetilised haigused võib jagada kolme klassi: monogeensed haigused, mida põhjustab mutatsioon ühes geenis, kromosomaalsed haigused, mis tulenevad kromosoomide või nende osade muutustest või puudumisest, ja polügeensed haigused, kus mutatsioone esineb mitmes geenis ning mille avaldumist võivad mõjutada ka elustiil ja keskkonnategurid (National Human Genome Research Institute, 2020). Töös käsitletavat kodade virvendusarütmiat ja laperdus on polügeensed haigused, mistõttu keskendutakse alljärgnevalt üksnes sellele klassile (Nielsen *et al.*, 2018).

Et hinnata polügeensete ehk paljudest geneetilistest muutustest mõjutatud komplekshaiguste esinemise riski, on vaja analüüsida paljusid väikese mõjuga geneetilisi variante ehk geeni DNA-järjestuses esinevaid püsivaid muutusi (Moorthie *et al.*, 2021). Üldjoontes võib geneetilised variandid jagada kahte klassi: ühenukleotiidsed variandid ja struktuurivariandid. Kõige levinum geneetiline varieeruvuse allikas inimese geenoomis on ühenukleotiidsed polümorfismid (*single nucleotide polymorphism*, SNP). (Frazer, Murray, Schork, Topol, 2009) Tegemist on DNA-järjestuse variatsioonidega, mis tekivad siis, kui üks nukleotiid erineb võrdlusjärjestusest, st on asendatud teisega (National Cancer Institute, *s. a.*). SNPde mõjud haigusriskile on enamasti aga üsna väikesed ja seega on nad eraldiseisvana madala prognoosivõimega. Seetõttu on mitme geneetilise variandi koosmõju hindamiseks võetud kasutusele polügeenne riskiskoor (*polygenic risk score*, PRS), mis kombineerib paljude SNPde mõjud. (Wray, Goddard, Visscher, 2007)

Polügeenne riskiskoor i -nda indiviidi jaoks on defineeritud kui k SNP kaalutud alleelidooside summa:

$$PRS_i = \sum_{j=1}^k w_j X_{ij},$$

kus X_{ij} on i -nda indiviidi j -nda SNP alleelidoos (efektialleelide arv) ning w_j on j -nda SNP kaal (Läll, 2019).

Alati pole selge, milline on optimaalne viis SNPde mõjude kombineerimiseks, et saavutada parim võimalik ennustuväärtus, kuna PRSi arvutamisel tuleb valida kaasatavate SNPde komplekt ja kaalud, mis pole üheselt määratud. Üks võimalus on tu-

letada need ülegenoomsest seoseuuringust (*genome-wide association study*, GWAS). (Läll, 2019) Ülegenoomses seoseuuringus otsitakse seoseid uuritava fenotüübi ja paljude SNPde vahel kogu genoomis. Nendest uuringutest saadud statistilised näitajad – efektisuurused ja p-väärtused – ongi aluseks SNPde valikule ja kaalumisele polügeensete riskiskooride koostamisel. (Choi, Mak, O'Reilly, 2020)

Kuna PRSid võimaldavad hinnata mitme geneetilise teguri panust haiguse kujunemisse, siis saab nende kaudu prognoosida inimese pärilikku eelsoodumust haiguse tekkeks (Polygenic Risk Score Task Force ..., 2021). Üksikule indiviidile omistatud PRSi väärtus ei ole aga otseselt interpreteeritav. PRSe tõlgendatakse tavaliselt võrdluses referentspopulatsiooni PRSi jaotusega, paigutades indiviidid nende riskiskoori alusel protsentiilivahemikesse. Selline lähenemine võimaldab jagada populatsiooni erinevatesse geneetilise riski rühmadesse (näiteks PRSi jaotuse ülemine 5%). (Torkamani, Wineinger, Topol, 2018) Oluline on silmas pidada, et PRS kajastab tõenäosuslikku, mitte absoluutset riski – kõrge PRS viitab küll suuremale haigestumise tõenäosusele, kuid ei määra haiguse vältimatut kujunemist, samas kui madal PRS ei välista haiguse tekkimise võimalust (National Human Genome Research Institute, 2020).

3 Elukestusanalüüs

Siinses peatükis kirjeldatav metoodika põhineb David Colletti (2014) raamatul „Modelling Survival Data in Medical Research“, kui ei ole märgitud teisiti.

Elukestusanalüüsi kasutatakse andmete puhul, mis kirjeldavad aega fikseeritud algmomendist kuni kindla sündmuse toimumiseni ehk lõppmomendini. Meditsiinilistes uuringutes on üldiselt algmomendiks indiviidi kaasamine uuringusse. Kui lõppmomendiks on surm, siis on tegemist tegeliku elukestuse uurimisega, kuid elukestusanalüüsi ei kasutata vaid elu kestuse analüüsimiseks – lõppmomendiks võib olla ka näiteks diagnoosi saamine, valu leevenemine või haiguse taasesinemine. Antud töös on algmomendiks aeg, mil indiviid liitus TÜ Eesti Geenivaramu kohordiga ja lõppmomendiks kodade virvendusarütmia või laperduse diagnoosi saamine.

Elukestusanalüüsi meetodite kasutamine on vajalik, sest kestusandmete ehk elulemusandmete puhul pole tavapäraste statistiliste meetodite rakendamine eelduste mitetäidetuse tõttu enamasti võimalik. Üks põhjus on see, et elukestused ei saa olla negatiivsed ja kestusandmed pole üldiselt sümmeetrilise jaotusega. Seetõttu ei saa ka eeldada, et andmed võiksid pärineda normaaljaotusest, mis on paljude standardsete statistiliste meetodite kasutamise aluseks. Peamine põhjus tuleneb aga andmete tsenseeritusest.

3.1 Tsenseeritus

Indiviidi elukestust nimetatakse tsenseerituks, kui lõppmomenti, see tähendab huvipakkuvat sündmust, pole vaadeldud. See võib tuleneda asjaolust, et uuring lõpeb enne huvipakkuva sündmuse toimumist (näiteks enne diagnoosi saamist) või juhul, kui isik on uuringust lahkunud või tema kohta pole alates mingist hetkest informatsiooni (näiteks isik kolib välismaale).

Oletame, et indiviid kaasatakse uuringusse hetkel t_0 . Olgu T juhuslik suurus, mis tähistab aega uuringuga liitumisest kuni huvipakkuva sündmuse toimumiseni ja C juhuslik suurus, mis tähistab aega uuringuga liitumisest kuni tsenseerimiseni. Kui indiviid on uurijate vaateväljas kuni hetkeni $t_0 + C$, siis nimetatakse hetke C

tsenseerimisajaks või tsenseeritud elukestuseks. Kui tegelik elukestus T on suurem kui tsenseeritud elukestus C ($C < T$) ehk indiviid kaob uurijate vaateväljast enne huvipakkuva sündmuse toimumist, on tegemist paremalt tsenseeritusega. Paremtsenseerituse korral on seega teada viimane ajahetk, mil indiviidiga polnud veel huvipakkuvat sündmust toimunud. Enamasti (ka siin töös analüüsitud andmetes) tekib paremalt tsenseerimine sellest, et uuring lõpetatakse enne kui huvipakkuv sündmus on kõigil toimunud. Seega on C aeg uuringuga liitumisest uuringu lõpuni.

3.1.1 Informatiivne ja mitteinformatiivne tsenseerimine

Tsenseeritud elukestusandmete analüüsimisel on oluline eeldus sõltumatu ehk mitteinformatiivne tsenseerimine ja järgnevates alapeatükkides kirjeldatavad standardsed elukestusanalüüsi meetodid kehtivad vaid sellel eeldusel. Mitteinformatiivne tsenseerimine tähendab, et tegelik elukestus T ja tsenseerimise aeg C on sõltumatud. Seega peavad tulevaste eluigade jaotused olema nii tsenseeritud kui mittetsenseeritud elukestustega indiviidide hulgas samad. Praktikas tähendab see, et uuringust lahkunud isikutel on pärast tsenseerimise hetke sama suur tõenäosus sündmuse esinemiseks kui neil, kes uuringusse alles jäävad.

Analüüsis kasutatavate geenivaramu andmete puhul on valdavalt tegemist mitteinformatiivse paremalt tsenseerimisega – tsenseerimine on tingitud asjaolust, et indiviid pole uuringu lõpuks diagnoosi saanud. Osa tsenseerimistest toimub aga ka seetõttu, et indiviid sureb enne huvipakkuva diagnoosi saamist muudel põhjustel. Seega on tegelikkuses tegemist informatiivse tsenseerimisega, kuna surm muudab huvipakkuva sündmuse (KVA diagnoosi) esinemise võimatuks. Selliseid sündmusi, mis välistavad uuritava sündmuse toimumise, nimetatakse konkureerivateks riskideks. Nende korrektseks käsitlemiseks kasutatakse konkureerivate riskide meetodeid, mis võimaldavad eristada eri sündmuste liike ja hinnata huvipakkuva sündmuse esinemist nende olemasolu arvestades.

3.2 Ajaskaala valik

Kestusandmete analüüsimisel on oluline valida sobiv ajaskaala, mis määrab muuhulgas selle, milliseid indiviide omavahel võrreldakse. Sõltuvalt ajaskaalast võib juhuslik suurus T väljendada kas kalendriaega või indiviidi vanust.

Kliiniliste uuringute kontekstis, kus indiviide jälgitakse üldiselt alates haiguse diagnoosimisest või ravi alustamisest kuni näiteks tervenemise, sümptomite leevenemise või surmani, on loomulik ajaskaala valik uuringus oldud aeg (*time-on-study*). Sel juhul võetakse inimese vanust arvesse, lisades see kovariaadina mudelisse. (Thiébaud, Bénichou, 2004)

Epidemioloogilistes uuringutes, kus huvi keskmes on haiguse esinemiseni kuluv aeg, eelistatakse sageli kasutada ajaskaalana vanust, kuna sel juhul ei ole uuringuga liitumise aeg üldjuhul seotud huvipakkuva haiguse esinemisega. Kui ajaskaalana kasutatakse vanust, siis võrreldakse igal ajahetkel uuritavaid teiste samaealiste indiviididega. Vanust ajaskaalana kasutades tuleb aga arvestada vasakult tõkestatusega, mis tähendab, et uuringusse jõuavad vaid need indiviidid, kes on uuringu alguses elus. Vaadeldes teatud vanust ei ole mõned indiviidid veel jälgimise all, samas kui mõned on juba vaatluse alt väljunud. Sellest tulenevalt ei muutu riskirühma suurus vanuse kasvades monotoonselt, erinevalt juhust, kui ajaskaalana kasutatakse uuringus oldud aega. (Thiébaud, Bénichou, 2004)

3.3 Üleelamis- ja riskifunktsioon

Elulemusandmete kirjeldamisel kasutatakse peamiselt kahte funktsiooni – üleelamisfunktsiooni ehk elulemusfunktsiooni ja riskifunktsiooni.

Olgu T juhuslik suurus, mis iseloomustab indiviidi elukestust ning võib omada mittemnegatiivseid väärtuseid ja $t \geq 0$ selle juhusliku suuruse realisatsioon. Olgu f juhuslikule suurusele T vastav tihedusfunktsioon. Sel juhul avaldub T jaotusfunktsioon kujul

$$F(t) = P(T \leq t) = \int_0^t f(s) ds \quad (1)$$

ja see iseloomustab tõenäosust, et elukestus on väiksem või võrdne ajaga t .

Üleelamisfunktsiooniks ehk elulemusfunktsiooniks nimetatakse tõenäosust, et elukestus on pikem kui aeg t , see tähendab, huvipakkuv sündmus ei toimu hetkeni t ja avaldub seega valemist (1) kujul

$$S(t) = P(T > t) = 1 - F(t).$$

Riskifunktsiooni kasutatakse väljendamaks mingi sündmuse riski toimuda hetkel t . Riskifunktsioon iseloomustab tõenäosust, et elukestus jääb lõpmatult väikesesse ajavahemikku $[t, t + \delta t)$, tingimusel, et elukestus ei ole väiksem kui t . Täpsemalt avaldub riskifunktsioon piirväärtusena antud tinglikust tõenäosusest ajaühiku kohta:

$$h(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T < t + \delta t \mid T \geq t)}{\delta t}. \quad (2)$$

3.4 Kaplan-Meieri hinnang

Kestusandmete analüüsimisel leitakse üldiselt esmalt üleelamisfunktsiooni Kaplan-Meieri hinnang. Oletame, et n indiviidi elukestuste seas on r unikaalset tsenseerimata väärtust, kus $r \leq n$. Moodustame nendest elukestustest variatsioonirea ning olgu j -nda elukestuse tähis $t_{(j)}$, $j = 1, 2, \dots, r$. Järjestatud ajahetked on seega $t_{(1)} < t_{(2)} < \dots < t_{(r)}$. Tähistame n_j -ga nende indiviidide arvu, kes on vahetult enne ajahetke $t_{(j)}$ vaatluse all (kaasa arvatud need, kelle elukestused on võrdsed ajaga $t_{(j)}$) ning märkigu d_j nende indiviidide arvu, kelle elukestused on võrdsed $t_{(j)}$ -ga, st kelle puhul toimus huvipakkuv sündmus hetkel $t_{(j)}$.

Kuna sündmused esinevad üksnes hetkedel $t_{(1)}, t_{(2)}, \dots, t_{(r)}$, piisab üleelamisfunktsiooni hindamiseks nende ajahetkede käsitlemisest. Vaatleme tõenäosusi

$$S(t_{(k)}) = P(T > t_{(k)}), \quad k = 1, 2, \dots, r,$$

mille kaudu on võimalik kirjeldada üleelamisfunktsiooni väärtust kõigil ajamomen-

tidel. Kasutades tõenäosuste korrutamise reeglit, saame

$$\begin{aligned}
P(T > t_{(k)}) &= P(T > t_{(k)} \mid T > t_{(k-1)}) \cdot P(T > t_{(k-1)}) \\
&= P(T > t_{(k)} \mid T > t_{(k-1)}) \cdot P(T > t_{(k-1)} \mid T > t_{(k-2)}) \cdot P(T > t_{(k-2)}) \\
&= \dots \\
&= \prod_{j=1}^k P(T > t_{(j)} \mid T > t_{(j-1)}),
\end{aligned} \tag{3}$$

kus kokkuleppeliselt $t_{(0)} = 0$. Paneme tähele, et iga korrutises oleva teguri korral kehtib

$$P(T > t_{(j)} \mid T > t_{(j-1)}) = 1 - P(t_{(j-1)} < T \leq t_{(j)} \mid T > t_{(j-1)}),$$

kus tinglik tõenäosus $P(t_{(j-1)} < T \leq t_{(j)} \mid T > t_{(j-1)})$ väljendab sündmuse toimumise tõenäosust ajavahemikus $(t_{(j-1)}, t_{(j)}]$, tingimusel, et indiviid on vahetult enne hetke $t_{(j)}$ vaatluse all. Selle tõenäosuse loomulik hinnang on $\frac{d_j}{n_j}$, seega saame, et

$$\hat{P}(T > t_{(j)} \mid T > t_{(j-1)}) = 1 - \frac{d_j}{n_j} = \frac{n_j - d_j}{n_j}.$$

Asendades need hinnangud korrutisse (3), saame Kaplan-Meieri hinnangu üleelamisfunktsioonile, mis avaldub kujul

$$\hat{S}(t) = \prod_{j=1}^k \left(\frac{n_j - d_j}{n_j} \right),$$

kus $t_{(k)} \leq t < t_{(k+1)}$, $k = 1, 2, \dots, r$ ja $t_{(r+1)} = \infty$. Kui vaadeldakse ajahetke enne esimese sündmuse toimumist, st $t < t_{(1)}$, siis on Kaplan-Meieri hinnang defineeritud kui $\hat{S}(t) = 1$. Kui suurima vaatluse puhul on tegemist tsenseeritud elukestusega t^* , siis ajahetkede $t > t^*$ jaoks pole hinnang $\hat{S}(t)$ defineeritud. Kui aga suurim vaatlus on tsenseerimata, siis $\hat{S}(t) = 0$, kui $t > t_{(r)}$. Kaplan-Meieri hinnang üleelamisfunktsioonile on seega astmefunktsioon, mille väärtus on sündmuste toimumishetkede vahel konstantne.

3.5 Konkureerivad riskid

Elukestusanalüüsi klassikalises käsitluses (konkureerivate riskide puudumisel) vaadeldakse indiviidi elulemust T kui aega uuringuga liitumisest kuni huvipakkuva sündmuse toimumiseni. Konkureerivaid riske arvesse võtvas analüüsis eristatakse riski hindamisel ka teisi sündmusi, mis huvipakkuva sündmuse toimumise välistavad. Sellisel juhul defineeritakse m juhuslikku suurust $T_1, T_2, \dots, T_m$, mis vastavad m võimaliku sündmuse toimumiste aegadele. Kuna korraga saab toimuda vaid üks sündmus, ei ole kõiki neist juhuslikest suurustest võimalik otseselt vaadelda – praktikas on sageli huvipakkuv (või ainsana vaadeldav) vaid esimesena toimuv sündmus. Seega on iga indiviidi puhul täheldatud elukestus $T = \min(T_1, T_2, \dots, T_m)$ ja sündmuse tüüp $P \in \{0, 1, \dots, m\}$, mis tähistab, milline sündmustest toimus (tsenseeritud vaatluste puhul, kus vaatlusperioodi lõpuks ei ole ühtegi sündmust toimunud, omandab P väärtuse 0).

Analüüsides ühe sündmuse riski, võttes konkureerivate riskidena arvesse ka teisi sündmusi, vaadeldakse selle sündmuse põhjusespetsiifilist esinemusfunktsiooni. Põhjusespetsiifiline esinemusfunktsioon iseloomustab tõenäosust, et sündmus s toimub ajahetkeks t , arvestades konkureerivate sündmuste võimalikkust, ja on defineeritud kui

$$F_s(t) = P(T \leq t, P = s),$$

kus $s = 1, 2, \dots, m$.

Sarnaselt tavalisele riskifunktsioonile $h(t)$, defineeritakse konkureerivate riskide korral iga sündmuse $s = 1, 2, \dots, m$ jaoks põhjusespetsiifiline riskifunktsioon

$$h_s(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T < t + \delta t, P = s \mid T \geq t)}{\delta t},$$

mis iseloomustab s -nda sündmuse riski toimuda hetkel t , arvestades konkureerivate sündmuste võimalikkust. Saab näidata, et

$$h_s(t) = \frac{f_s(t)}{S(t)},$$

kus f_s on põhjusespetsiifiline tihedusfunktsioon, ja seega saab põhjusespetsiifilise esinemusfunktsiooni avaldada kui

$$F_s(t) = \int_0^t h_s(u)S(u) du. \quad (4)$$

3.5.1 Aalen-Johanseni hinnang

Kui konkureerivaid riske ei võeta arvesse (st surma käsitletakse tsenseerimisena), saadakse nihkega hinnangud – enamasti on Kaplan-Meieri hinnangud ülehinnatud. Aalen-Johanseni hinnang on Kaplan-Meieri hinnangu üldistus konkureerivate riskide korral ja võimaldab hinnata põhjusespetsiifilist esinemusfunktsiooni.

Oletame, et n indiviidi elukestuste seas on r unikaalset tsenseerimata väärtust ja moodustame nendest elukestustest variatsioonirea $t_{(1)} < t_{(2)} < \dots < t_{(r)}$, kus sündmuste toimumisaegade hulgas on arvestatud kõiki m tüüpi sündmusi. Tähistagu n_j vahetult enne ajahetke $t_{(j)}$ vaatluse all olevate indiviidide arvu ja märkigu d_{sj} indiviidide arvu, kelle puhul toimus sündmus s hetkel $t_{(j)}$. Seega $d_j = \sum_{s=1}^m d_{sj}$ tähistab nende indiviidide arvu, kellel toimus ükskõik milline sündmustest hetkel $t_{(j)}$.

Vaadeldes diskreetset versiooni põhjusespetsiifilisest riskifunktsioonist, saab näidata, et $h_s(t_{(j)}) \approx \frac{d_{sj}}{n_j}$. Asendades selle hinnangu avaldisse (4), saame Aalen-Johanseni hinnangu põhjusespetsiifilisele esinemusfunktsioonile:

$$\hat{F}_s(t) = \sum_{j:t_{(j)} \leq t} \hat{S}(t_{(j-1)}) \frac{d_{sj}}{n_j},$$

kus $\hat{S}(t_{(j-1)})$ on Kaplan-Meieri hinnang üldisele elulemusfunktsioonile hetkel $t_{(j-1)}$.

3.6 Coxi võrdeliste riskide mudel

Elukestusanalüüsis on üldjuhul fookuses kirjeldavate tunnuste mõju uurimine elulemusele. Üldiselt soovitakse seega luua mudelit riskifunktsioonile (2) ja selleks kasutatakse peamiselt võrdeliste riskide mudeleid.

Olgu vaatluse all n indiviidi ja olgu iga indiviidi puhul mõõdetud p kirjeldavat tun-

nust $X_1, \dots, X_p$. Tähistame kirjeldavate tunnuste väärtuste vektori $\mathbf{x} = (x_1, \dots, x_p)'$. Olgu i -nda indiviidi, $i = 1, 2, \dots, n$, riskifunktsioon $h_i(t)$ ja kirjeldavate tunnuste väärtuste vektor $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})'$. Lisaks tähistame $h_0(t)$ -ga riskifunktsiooni indiviidi jaoks, kelle puhul on kõigi kirjeldavate tunnuste väärtused võrdsed nulliga, st $\mathbf{x} = \vec{0}$. Funktsiooni $h_0(t)$ nimetatakse baasriskifunktsiooniks.

Võrdeliste riskide mudeli korral avaldub i -nda indiviidi riskifunktsioon kujul

$$h_i(t) = \psi(\mathbf{x}_i)h_0(t),$$

kus $\psi(\mathbf{x}_i) \geq 0$ on i -nda indiviidi kirjeldavate tunnuste väärtuste vektori $\mathbf{x}_i$ funktsioon, mis kirjeldab riskimäärade suhet (*hazard ratio*, HR). Funktsiooni $\psi(\mathbf{x}_i)$ enimlevinud kuju on $\psi(\mathbf{x}_i) = e^{\eta_i}$, kus η_i on vektori $\mathbf{x}_i$ lineaarkombinatsioon

$$\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} = \boldsymbol{\beta}' \mathbf{x}_i$$

ning $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ on vektor kirjeldavate tunnuste kordajatest. Võrdeliste riskide mudeli saab seega kirja panna kujul

$$h_i(t) = e^{\boldsymbol{\beta}' \mathbf{x}_i} h_0(t) = e^{\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}} h_0(t). \quad (5)$$

Võrdusest (5) järeldub, et kui ühe kirjeldava tunnuse X_k ($k = 1, 2, \dots, p$) väärtus x_k suureneb ühe ühiku võrra, siis $h_i(t)$ muutub e^{β_k} korda:

$$\frac{h_i(t; x_1, \dots, x_k + 1, \dots, x_p)}{h_i(t; x_1, \dots, x_k, \dots, x_p)} = e^{\beta_k}.$$

Kui mudeli puhul pole tehtud eeldusi baasriskifunktsiooni kuju kohta, siis on tegemist Coxi võrdeliste riskide mudeliga ja sel juhul jäetakse $h_0(t)$ kuju määramata ning hinnatakse vaid parameetreid $\beta_1, \dots, \beta_p$, see tähendab, tegemist on poolparameetrilise mudeliga. Parameetrilisel juhul on mudelis baasriskifunktsioon $h_0(t)$ määratud vastava jaotuse parameetritega.

Mudeli parameetrite $\beta_1, \dots, \beta_p$ hindamiseks kasutatakse osalise tõepära funktsiooni. Eeldame, et uuringus on n indiviidi, kelle seas on vaadeldud r sündmust. Tähistame

nende sündmuste järjestatud toimumiste ajad $t_{(1)} < t_{(2)} < \dots < t_{(r)}$, kus $t_{(j)}$ on j -nda sündmuse toimumise aeg. Olgu $R(t_{(j)})$ nende indiviidide hulk, kelle puhul pole vahetult enne hetke $t_{(j)}$ sündmust vaadeldud ning keda pole tsenseeritud. Seda hulka nimetatakse riskigrupiks. Sel juhul avaldub osalise tõepära funktsioon Coxi võrdeliste riskide mudelile kujul

$$L(\boldsymbol{\beta}) = \prod_{j=1}^r \frac{e^{\boldsymbol{\beta}' \mathbf{x}_{(j)}}}{\sum_{l \in R(t_{(j)})} e^{\boldsymbol{\beta}' \mathbf{x}_l}},$$

kus $\mathbf{x}_{(j)}$ on kirjeldavate tunnuste vektor indiviidi jaoks, kelle puhul toimus huvipakkuv sündmus j -ndal ajahetkel $t_{(j)}$. Hinnangud mudeli (5) parameetritele on seega parameetrite $\boldsymbol{\beta}$ väärtused, mis maksimeerivad funktsiooni $L(\boldsymbol{\beta})$.

3.7 Võrdeliste riskide eelduse kontrollimine

Coxi võrdeliste riskide mudeli keskne eeldus on, et riskimäärad on võrdelised ja riskimäärade suhe on ajas konstantne. Selle eelduse paikapidavuse kontrollimiseks kasutatakse kaalutud Schoenfeldi jääkide analüüsi.

Olgu vaadeldud n indiviidi elukestusi, kellest r subjektile on vaadeldud huvipakkuv sündmus ning $n - r$ subjekti on paremalt tsenseeritud. Oletame, et elukestustele on sobitatud Coxi võrdeliste riskide mudel, mis sisaldab p kirjeldavat tunnust $X_1, \dots, X_p$. Siis i -ndale indiviidile ($i = 1, 2, \dots, n$) sobitatud riskifunktsioon on

$$\hat{h}_i(t) = e^{\hat{\boldsymbol{\beta}}' \mathbf{x}_i} \hat{h}_0(t),$$

kus $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)'$ on mudeli parameetrite hinnangute vektor ja $\hat{h}_0(t)$ on baasriskifunktsiooni hinnang.

Esmalt defineeritakse Schoenfeldi jäägid, mille puhul erinevalt tavapärastest jääkidest, ei ole iga indiviidi kohta üksainus jäägi väärtus, vaid väärtuste kogum – üks iga kaasatud selgitava muutuja kohta sobitatud Coxi mudelis. Olgu δ_i sündmuse indikaatortunnus, mille väärtus on 1, kui i -nda indiviidi puhul toimus huvipakkuv sündmus, st elukestus ei ole tsenseeritud, ning 0 vastasel juhul (i -nda indiviidi elu-

kestus on tsenseeritud). Schoenfeldi jääk i -nda indiviidi ja j -nda kirjeldava tunnuse X_j korral avaldub kujul

$$r_{Sji} = \delta_i \{x_{ji} - \hat{a}_{ji}\},$$

kus x_{ji} on j -nda kirjeldava tunnuse väärtus i -nda indiviidi puhul,

$$\hat{a}_{ji} = \frac{\sum_{l \in R(t_i)} x_{jl} e^{\hat{\beta}' \mathbf{x}_l}}{\sum_{l \in R(t_i)} e^{\hat{\beta}' \mathbf{x}_l}}$$

ning $R(t_i)$ on riskigrupp ajahetkel t_i .

Edasi defineeritakse kaalutud Schoenfeldi jäägid. Tähistame i -nda indiviidi Schoenfeldi jääkide vektori $\mathbf{r}_{Si} = (r_{S1i}, r_{S2i}, \dots, r_{Spi})'$. Kaalutud Schoenfeldi jäägid i -nda indiviidi jaoks on vektori

$$\mathbf{r}_{Si}^* = r \cdot \text{var}(\hat{\beta}) \mathbf{r}_{Si}$$

komponendid, kus r tähistab n indiviidi hulgas toimunud sündmuste arvu ja $\text{var}(\hat{\beta})$ on sobitatud Coxi mudeli hinnatud parameetrite kovariatsioonimaatriks.

Kaalutud Schoenfeldi jääkide korral kehtib omadus

$$E(r_{Sji}^*) \approx \beta_j(t_i) - \hat{\beta}_j,$$

kus r_{Sji}^* tähistab i -nda kaalutud Schoenfeldi jäägi väärtust j -nda kirjeldava tunnuse X_j korral, $\beta_j(t_i)$ on kovariaadi X_j ajas muutuva kordaja väärtus ajahetkel t_i ja $\hat{\beta}_j$ on j -nda argumenttunnuse parameetri β_j hinnang. Väärtuste $r_{Sji}^* + \hat{\beta}_j$ kujutamine graafikul annab teavet ajas muutuva kordaja $\beta_j(t)$ kuju kohta, mis võimaldab kontrollida võrdeliste riskide eelduse kehtivust. Nende väärtuste paiknemine horisontaalsel joonel viitab sellele, et tunnuse X_j kordaja on ajas konstantne ja võrdeliste riskide eeldus on täidetud.

3.8 Kihitatud Coxi võrdeliste riskide mudel

Võib esineda olukordi, kus võrdeliste riskide eeldus ei kehti kogu andmestiku lõikes, kuid on rahuldatud alamrühmades (kihtides). Sel juhul kasutatakse sageli kihitatud

Coxi võrdeliste riskide mudelit, mille puhul on riskil lubatud kihtide lõikes erineda, st baasriskifunktsioonid on kihiti erinevad. Igas kihis eeldatakse siiski võrdeliste riskide kehtivust kõigi argumenttunnuste korral ja hinnatavad parameetrid β on kõigis kihtides samad. Tunnuseid, mille alusel kihitamine toimub, ei lisata mudelisse argumenttunnustena.

Oletame, et vaatleme k kihti ja j -ndasse kihti kuulub n_j indiviidi. Tähistame j -nda kihi baasriskifunktsiooni $h_{0j}(t)$, $j = 1, 2, \dots, k$, ja olgu $h_{ij}(t)$ j -ndasse kihti kuuluva i -nda indiviidi riskifunktsioon, $i = 1, 2, \dots, n_j$. Siis kihitatud Coxi võrdeliste riskide mudel p argumenttunnusega avaldub kujul

$$h_{ij}(t) = e^{\beta' \mathbf{x}_{ij}} h_{0j}(t),$$

kus $\mathbf{x}_{ij}$ on j -ndasse kihti kuuluva i -nda indiviidi kirjeldavate tunnuste väärtuste vektor.

Mudeli parameetrid $\beta_1, \dots, \beta_p$ hinnatakse sarnaselt tavalisele Coxi võrdeliste riskide mudelile maksimeerides osalise tõepära funktsiooni, mis avaldub kõikide kihtide osalise tõepära funktsioonide korrutisena (Kleinbaum, Klein, 1997).

3.9 Harrelli C-indeks

Harrelli C-indeks (*concordance index*) on elukestusanalüüsis sageli kasutatav mõõdik, mille abil saab hinnata mudeli diskrimineerimisvõimet ehk seda, kui korrektselt suudab mudel indiviide nende riski põhjal järjestada. C-indeks on defineeritud kui osakaal kõigist võrreldavatest indiviidide paaridest, mille puhul on mudeli ennustused ja tegelikud tulemused kooskõlas. Seega võrreldakse C-indeksi arvutamisel iga paari puhul ennustatud riskiskooride (st mudeli lineaarse osa hinnangute) järjestust vaadeldud sündmuste toimumisaegade järjestusega. (Harrell, Lee, Mark, 1996)

Indiviidide paar loetakse võrreldavaks, kui nende elukestused erinevad ja lühema elukestusega indiviidi puhul on sündmus toimunud (elukestus on tsenseerimata). Kui mõlema indiviidi elukestused on tsenseeritud või võrdsed, siis ei ole paar võrreldav. (Harrell, Lee, Mark, 1996) Indiviidide paari nimetatakse kooskõlaliseks, kui sub-

jektile, kes kogeb sündmust vastavalt kasutatavale ajaskaalale varem, omistatakse kõrgem ennustatud riskiskoor kui subjektile, kellel sündmus esineb hiljem või kelle puhul pole vaatlusperioodi jooksul sündmust toimunud. Vastasel juhul nimetatakse paari ebakooskõlaliseks. Kui mudeli poolt hinnatud riskiskoorid on mõlemal indiviidil võrdsed, kuid sündmuste toimumiste ajad erinevad, siis käsitletakse sellist paari osaliselt kooskõlalisena. (Therneau, Atkinson, 2026)

Tähistagu c , d ja t_x vastavalt kooskõlaliste, ebakooskõlaliste ja osaliselt kooskõlaliste paaride arvu. C-indeks on kooskõla tõenäosuse hinnang, mis on defineeritud kui tõenäosus, et kahel juhuslikult valitud indiviidil on riskiennustused õigesti järjestatud (Therneau, Atkinson, 2026):

$$C = \hat{P}(\text{kooskõla}) = \frac{c + t_x/2}{c + d + t_x}.$$

C-indeksi väärtused jäävad vahemikku $[0, 1]$, kus väärtus 1 viitab täiuslikule diskrimineerimisvõimele, st mudeli ennustused langevad täielikult kokku tegeliku sündmuste toimumise järjekorraga, ning väärtus 0 tähendab, et mudel järjestab kõik indiviidide paarid vastupidiselt. Väärtuse 0,5 korral prognoosib mudel juhuslikult. (Harrell, Lee, Mark, 1996)

Ülevaade käesolevas peatükis kirjeldatud elukestusanalüüsi meetodite rakendamisest tarkvaras R on esitatud lisas 1.

4 Logistiline regressioonimudel

Logistiline regressioon on statistiline meetod, mida kasutatakse selgitavate tunnuste ja binaarse tulemuse esinemise tõenäosuse vahelise seose modelleerimiseks. Olgu va-
limis n vaatlust ja p seletavat tunnust $X_1, \dots, X_p$. Olgu Y_i ($i = 1, 2, \dots, n$) binaarne
juhuslik suurus, mis kirjeldab huvipakkuva sündmuse toimumist i -ndal indiviidil.
Tavaliselt kodeeritakse uuritava tunnuse väärtused 1/0, kus 1 tähistab sündmuse
esinemist. Olgu $p_i = P(Y_i = 1)$ sündmuse toimumise tõenäosus i -ndal indiviidil.
(Hosmer, Lemeshow, Sturdivant, 2013)

Binaarse uuritava tunnuse korral kasutatakse logit-seosefunktsiooni

$$\text{logit}(p_i) = \ln \frac{p_i}{1 - p_i},$$

kus $\frac{p_i}{1-p_i}$ tähistab sündmuse toimumise šanssi, mis näitab mitu korda on huvipakkuva
sündmuse toimumine tõenäolisem selle mittetoimumisest. Logistilise regressiooni
mudeliga hinnatakse argumenttunnuste lineaarkombinatsiooni kaudu šansi logaritmi

$$\ln \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi},$$

kus β_0 on vabaliige, $\beta_1, \beta_2, \dots, \beta_p$ on regressioonikordajad ning $x_{1i}, x_{2i}, \dots, x_{pi}$ on
 i -nda indiviidi selgitavate tunnuste väärtused. Mudeli tundmatud parameetrid hin-
natakse suurima tõepära meetodil. (Hosmer, Lemeshow, Sturdivant, 2013)

Regressioonikordajate statistilise olulisuse hindamiseks kasutatakse Wald'i z-statisti-
kut. Iga parameetri β_j ($j = 1, 2, \dots, p$) korral kontrollitakse kahepoolse testi abil
hüpoteesi $H_0 : \beta_j = 0$. Z-statistik arvutatakse hinnatud kordaja ja selle standardvea
suhtena

$$z = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)},$$

kus $\hat{\beta}_j$ on hinnatud regressioonikordaja ning $\text{se}(\hat{\beta}_j)$ on selle standardviga. Mida suu-
rem on z-statistiku absoluutväärtus, seda tugevam on statistiline tõendus, et vastav
regressioonikordaja erineb nullist, st selgitav tunnus on seotud uuritava sündmuse
esinemisega. (Hosmer, Lemeshow, Sturdivant, 2013)

5 Andmete analüüs

5.1 Ülevaade andmetest

TÜ Eesti Geenivaramu sisaldab põhjalikku teavet nii geenidoonorite geeni- kui ka terviseandmete kohta. Kõikidele doonoritele on tehtud genotüpiseerimine, mis kaardistab üle 700 000 SNP markeri, sealhulgas Eesti populatsiooni-spetsiifilisi variante. Geenivaramusse kogutud terviseandmed pärinevad nii doonorite endi poolt täidetud küsimustikest, mis hõlmavad muu hulgas elustiili ja haridusega seotud teavet, kui ka terviseandmeid sisaldavatest registritest. (TÜ Genoomika Instituut, *s. a.*) Geenidoonorite värbamine toimus kahes peamises laines: varasem kohort (enne 2018. aastat liitunud) kaasati valdavalt perearstide ja värbamispunktide kaudu, samas kui alates 2018. aastast toimunud riikliku värbamiskampaania käigus lihtsustus liitumisprotsess märkimisväärselt, mille tõttu erinevad need kaks kohorti kaasamisviisi poolest (Milani *et al.*, 2025).

Andmestik hõlmas 211 183 geenidoonori andmeid, neist 73 025 olid mehed (34,6%) ja 138 158 naised (65,4%). Kodade virvendusarütmia või laperduse diagnoosiga indiviide oli 16 008 ehk ligikaudu 7,6% kõigist doonoritest ja nad jagunesid kahte rühma: nendeks, kellel oli KVA diagnoositud enne TÜ Eesti Geenivaramuga liitumist, ning nendeks, kes said diagnoosi peale liitumist. Enne geenivaramuga liitumist KVA diagnoosi saanud oli 7467 (3,5% kõigist doonoritest, 46,6% KVA diagnoosiga indiviididest), samas kui neid, kes said diagnoosi peale geenivaramuga liitumist, oli 8541 (4,0% kõigist doonoritest, 53,4% KVA diagnoosiga indiviididest).

Elukestusanalüüsis kasutatava polügeense riskiskoori valimiseks kasutati nende indiviidide andmeid, kellel oli diagnoositud kodade virvendusarütmia või laperdus enne TÜ Eesti Geenivaramuga liitumist või kes ei saanud üldse diagnoosi. Elukestusanalüüsi läbiviimisel seevastu vaadeldi vaid neid indiviide, kellel ei olnud diagnoositud kodade virvendusarütmia või laperdust enne geenivaramuga liitumist.

5.2 PRSi valimine

Kõigi doonorite geeniproovidele on arvatatud 47 KVA polügeenset riskiskoori, kasutades PGS kataloogis avaldatud andmeid (Polygenic Score (PGS) Catalog, *s. a.*). Skoorid pärinevad uuringutest, mille valimid koosnesid valdavalt Euroopast pärit isikutest ja ei hõlmanud TÜ Eesti Geenivaramu andmeid. Elukestusanalüüsis kasutatava PRSi valimiseks hinnati iga skoori puhul logistilise regressiooni mudel, milles sõltuv tunnus oli KVA diagnoosi olemasolu enne geenivaramuga liitumist ning selektavad tunnused skaleeritud PRS, sugu ja vanus geenivaramuga liitumisel. Kasutatav PRS valiti 47 kandidaadi hulgast kõrgeima z-statistiku alusel.

Tabelis 1 on esitatud kümme suurima z-statistiku väärtusega riskiskoori. Iga skoori kohta on välja toodud ka mudelis hinnatud kordaja $\hat{\beta}$ väärtus ja selle standardviga $se(\hat{\beta})$.

Tabel 1: 10 suurima z-statistiku väärtusega riskiskoori

PRS	$\hat{\beta}$ (se)	z-statistik
PGS005071	0,566 (0,013)	43,129
PGS005072	0,562 (0,013)	42,943
PGS005067	0,566 (0,013)	42,657
PGS005070	0,549 (0,013)	42,306
PGS005065	0,550 (0,013)	42,208
PGS002756	0,550 (0,013)	42,000
PGS005063	0,523 (0,013)	40,541
PGS005069	0,504 (0,013)	38,857
PGS000331	0,490 (0,013)	38,696
PGS005061	0,495 (0,013)	37,690

Töös on edaspidi kasutatud polügeenset riskiskoori PGS005071, mis pärineb Gunn *et al.* (2024) uuringust „Comparison of methods for building polygenic scores for diverse populations“. Samas uuringus on välja töötatud veel seitse parimate skooride seas olevat PRSi.

5.3 Kirjeldav analüüs

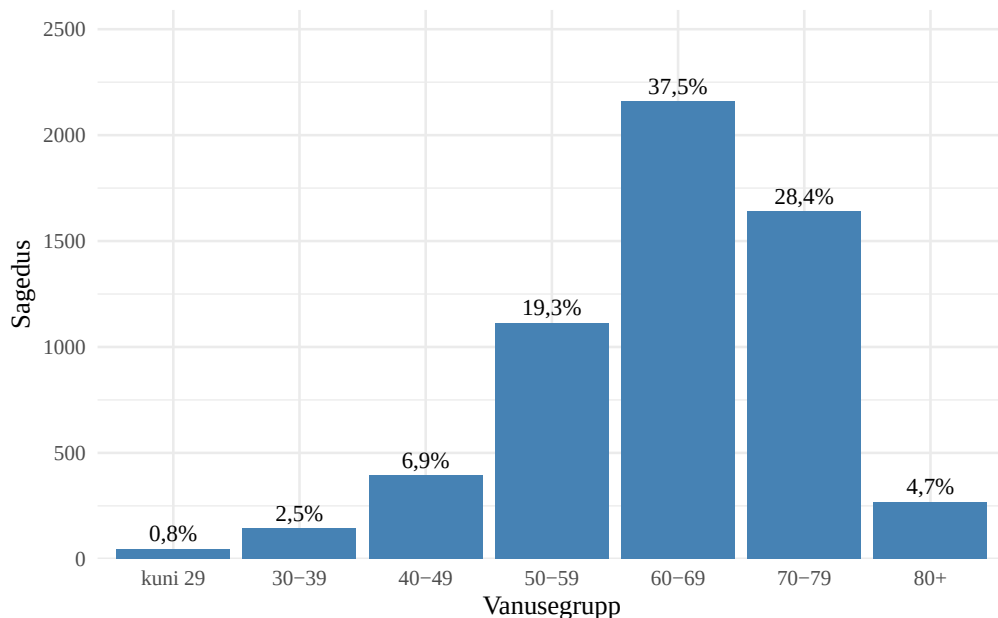
Edasise analüüsi läbiviimisel kaasati need indiviidid, kellel ei olnud diagnoositud kodade virvendusarütmiaid või laperdust enne TÜ Eesti Geenivaramuga liitumist. Vaadeldi vaid neid indiviide, kes ei olnud liitumise hetkel vanemad kui 70 aastat, sest vanemaealistel võivad tõenäolisemalt esineda konkureerivad riskid. Lisaks jäeti analüüsist välja indiviidid, kellel esines puuduvaid väärtusi huvipakkuvates (mudelisse kaasatavates) tunnustes.

Kodade virvendusarütmia ja laperduse potentsiaalsete riskifaktoritena kaasati töös lisaks geneetilisele riskiskoorile järgmised tunnused: kõrgeim omandatud haridustase, suitsetamise staatus, kehamassiindeks (KMI) ning mitmed enne geenivaramuga liitumist registreeritud diagnoosid, sealhulgas kilpnäärme haigusseisundid, diabeet, kõrgvererõhkhaigused, südame isheemiatõved ning unehäired. Haridustasemel on kolm kategooriat: alla keskhariduse, keskharidus ja kõrgharidus. Suitsetamise staatus jaguneb kaheks: mitesuitsetajad ja suitsetajad (liitumishetke seisuga). Järgnevalt on välja toodud kaasatud diagnoosid koos vastavate RHK-10 koodidega:

- E05 – hüpertüreoos,
- E10 – insuliinisõltuv suhkurtõbi (I tüüpi diabeet),
- E11 – insuliinisõltumatu suhkurtõbi (II tüüpi diabeet),
- I10–I15 – kõrgvererõhkhaigused,
- I21 – äge müokardiinfarkt,
- I22 – korduv müokardiinfarkt,
- I25 – krooniline südame isheemiatõbi,
- I50 – südamepuudulikkus,
- G47 – orgaanilised unehäired.

Lõplik valim koosnes 171 106 indiviidist, kellest 5764-l (3,4%) diagnoositi uuringu perioodi jooksul KVA. Valimis oli 113 013 naist (66,0%) ning 58 093 meest (34,0%). KVA diagnoosi saanud isikute osakaal oli meeste seas suurem kui naistel, vastavalt 4,3% ja 2,9%. Keskmise vanus geenivaramuga liitumisel oli ligikaudu 41,5 aastat ja keskmiselt olid subjektid vaatluse all 9,0 aastat. Keskmise diagnoosi saamise vanus oli 64,3 aastat, meeste puhul 62,2 aastat ja naiste puhul 65,9 aastat.

Joonisel 1 on kujutatud diagnooside arvud ja vastavad osakaalud vanusegrupiti, kus vanus tähistab vanust diagnoosi saamisel. Enamik KVA diagnoose (85,2%) esines vanuses 50–79 aastat. Kõige rohkem on indiviide, kes on diagnoosi saanud vanuses 60–69 aastat, järgnevad vanusegrupid 70–79 ja 50–59 aastat.



Joonis 1: Diagnooside arvud vastavalt diagnoosi saamise vanusele

Edasises kirjeldavas analüüsis on esitatud jaotused kohortide, liitumise vanusegrupi ja soo lõikes. Kohorte on eristatud lähtuvalt geenivaramu doonorite värbamise kahest peamisest lainest, mille tõttu võivad liitumise põhjused ja seega ka kohortide koosseis mõnevõrra erineda. Vanusegrupi ja soo eristamine võimaldab omakorda hinnata erinevusi kaasatavate riskifaktorite esinemissageduses, kuna on alust eeldada, et mitmete tervisenäitajate ja riskitegurite mõju võib nendes rühmades erineda.

Tabelis 2 on esitatud 2002.–2017. aastal geenivaramuga liitunud doonorite kirjeldavad statistikud soo ja liitumise vanuserühma lõikes. Esimeses kohordis oli 45 803 indiviidi, kellest suurema osa moodustasid naised vanuserühmas 18–59 aastat (57,1%). Vanuses 60+ indiviidide osakaal oli suhteliselt väike (13,3%), kuid samas esines neil kõige suurema tõenäosusega liitumisjärgselt KVA diagnoos (32,8% meestest ja 29,2% naistest). Mõlemas vanuserühmas oli KVA diagnoosi esinemissagedus suurem meeste seas.

Tabel 2: Esimese kohordi (2002–2017 liitunud) kirjeldavad statistikud soo ja vanuserühma lõikes

Liitunud 2002–2017	18–59		60+	
	Mehed	Naised	Mehed	Naised
N	13 575	26 155	1 984	4 089
Suitsetajad	6 138 (45,2%)	7 088 (27,1%)	518 (26,1%)	419 (10,2%)
Haridus				
Alla keskhariduse	1 451 (10,7%)	1 457 (5,6%)	475 (23,9%)	725 (17,7%)
Keskharidus	7 790 (57,4%)	13 056 (49,9%)	949 (47,8%)	2 357 (57,6%)
Kõrgharidus	4 334 (31,9%)	11 642 (44,5%)	560 (28,2%)	1 007 (24,6%)
KMI, keskmine (sd)	26,0 (4,6)	25,4 (5,4)	28,0 (4,5)	29,3 (5,4)
Liitumishetkeks diagnoositud haigused				
Hüpertüreoos	70 (0,5%)	790 (3,0%)	16 (0,8%)	180 (4,4%)
I tüüpi diabeet	187 (1,4%)	200 (0,8%)	76 (3,8%)	102 (2,5%)
II tüüpi diabeet	485 (3,6%)	727 (2,8%)	313 (15,8%)	619 (15,1%)
Äge või korduv müokardiinfarkt	300 (2,2%)	82 (0,3%)	226 (11,4%)	118 (2,9%)
Krooniline südame isheemiatõbi	443 (3,3%)	397 (1,5%)	505 (25,5%)	661 (16,2%)
Südamepuudulikkus	422 (3,1%)	582 (2,2%)	466 (23,5%)	713 (17,4%)
Orgaanilised unehäired	604 (4,4%)	1 161 (4,4%)	168 (8,5%)	455 (11,1%)
Kõrgvererõhkhaigused	3 314 (24,4%)	4 751 (18,2%)	1 331 (67,1%)	2 804 (68,6%)
KVA diagnoos liitumisjärgselt	979 (7,2%)	1 098 (4,2%)	651 (32,8%)	1 196 (29,2%)

Vähemalt keskharidusega inimeste osakaal oli naiste seas kõrgem kui meestel ning suitsetamine oli märgatavalt levinum meeste ja nooremaealiste seas. Vanemas vanuserühmas oli keskmine kehamassiindeks kõrgem kui noorematel. Kõigi kaasuvate hai-

guste esinemissagedus suureneb vanuse kasvades, see kehtib eriti II tüüpi diabeedi, kroonilise südame isheemiatõve, südamepuudulikkuse ja kõrgvererõhkhaiguste puhul. Pea kõigi kaasuvate haiguste puhul oli diagnooside osakaal kõrgem meeste seas, seda sõltumata vanusegrupist.

Tabelis 3 on esitatud 2018.–2021. aastal geenivaramuga liitunud doonorite kirjeldavad statistikud soo ja liitumise vanuserühma lõikes. Teine kohort koosnes 125 303 indiviidist ja sarnaselt esimesele kohordile oli enim 18–59-aastaselt liitunud naised (57,0%).

Tabel 3: Teise kohordi (2018–2021 liitunud) kirjeldavad statistikud soo ja vanuserühma lõikes

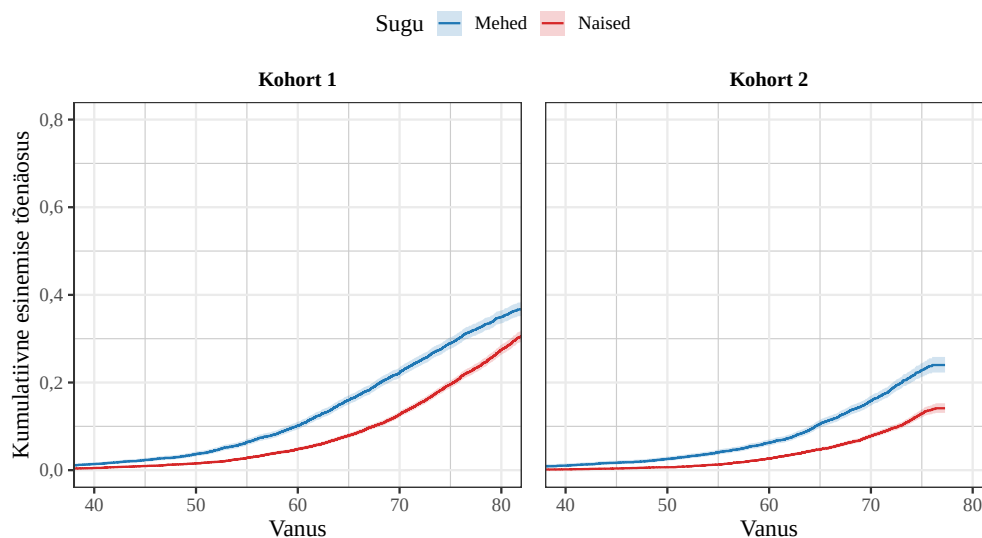
Liitunud 2018–2021	18–59		60+	
	Mehed	Naised	Mehed	Naised
N	37 820	71 448	4 714	11 321
Suitsetajad	10 330 (27,3%)	13 505 (18,9%)	820 (17,4%)	1 422 (12,6%)
Haridus				
Alla keskhariduse	2 774 (7,3%)	2 538 (3,6%)	351 (7,4%)	500 (4,4%)
Keskharidus	18 355 (48,5%)	26 959 (37,7%)	2 377 (50,4%)	6 244 (55,2%)
Kõrgharidus	16 691 (44,1%)	41 951 (58,7%)	1 986 (42,1%)	4 577 (40,4%)
KMI, keskmine (sd)	26,7 (4,4)	25,0 (5,2)	28,3 (4,4)	28,0 (5,2)
Liitumishetkeks diagnoositud haigused				
Hüpertüreoos	252 (0,7%)	2 622 (3,7%)	62 (1,3%)	628 (5,5%)
I tüüpi diabeet	332 (0,9%)	422 (0,6%)	102 (2,2%)	139 (1,2%)
II tüüpi diabeet	1 029 (2,7%)	1 403 (2,0%)	759 (16,1%)	1 281 (11,3%)
Äge või korduv müokardiinfarkt	246 (0,7%)	72 (0,1%)	243 (5,2%)	118 (1,0%)
Krooniline südame isheemiatõbi	617 (1,6%)	457 (0,6%)	775 (16,4%)	1 029 (9,1%)
Südamepuudulikkus	387 (1,0%)	602 (0,8%)	421 (8,9%)	839 (7,4%)
Orgaanilised unehäired	4 065 (10,7%)	6 575 (9,2%)	963 (20,4%)	2 698 (23,8%)
Kõrgvererõhkhaigused	8 903 (23,5%)	9 274 (13,0%)	2 970 (63,0%)	6 464 (57,1%)
KVA diagnoos liitumisjärgselt	455 (1,2%)	401 (0,6%)	434 (9,2%)	550 (4,9%)

Ka teises kohordis oli KVA diagnoosi olemasolu sagedasem vanemas vanuserühmas ja meeste seas. Küll aga oli KVA diagnoosi esinemissagedus liitumisjärgselt märga-

tavalt madalam kui esimeses kohordis, kuid see võib olla osaliselt seotud ka lühema jälgimisajaga. Vähemalt keskharidusega isikute osakaal oli ka teises kohordis naiste seas kõrgem kui meestel, kuid võrreldes esimese kohordiga on vähemalt keskharitute osakaal kõigis gruppides suurem. Sarnaselt esimesele kohordile oli suitsetajaid rohkem meeste seas ja nooremas vanuserühmas, samas on suitsetamise levimus võrreldes esimese kohordiga pea kõigis gruppides madalam. Ka teises kohordis suureneb kõigi kaasuvate haiguste esinemissagedus vanuse kasvades, ent võrreldes esimese kohordiga olid teises kohordis kaasuvate haiguste esinemissagedused üldjuhul madalamad. Küll aga oli teises kohordis orgaanilise unehäire diagnoosiga indiviidide osakaal suurem. Pea kõigi kaasuvate haiguste puhul oli meeste seas sõltumata vanusegrupist diagnooside osakaal kõrgem.

5.4 Kodade virvendusarütmia ja laperduse riskitegurid

Esmalt vaadeldi KVAse haigestumist soo ja kohordi lõikes. Joonisel 2 on kujutatud Aalen-Johanseni hinnangud KVA esinemusfunktsioonile koos 95% usaldusintervallidega. Hinnangute leidmisel on käsitletud surma konkureeriva riskina, ajaskaalana on kasutatud vanust ja on arvestatud vasakult tõkestatuse ning paremalt tsenseeritusega.



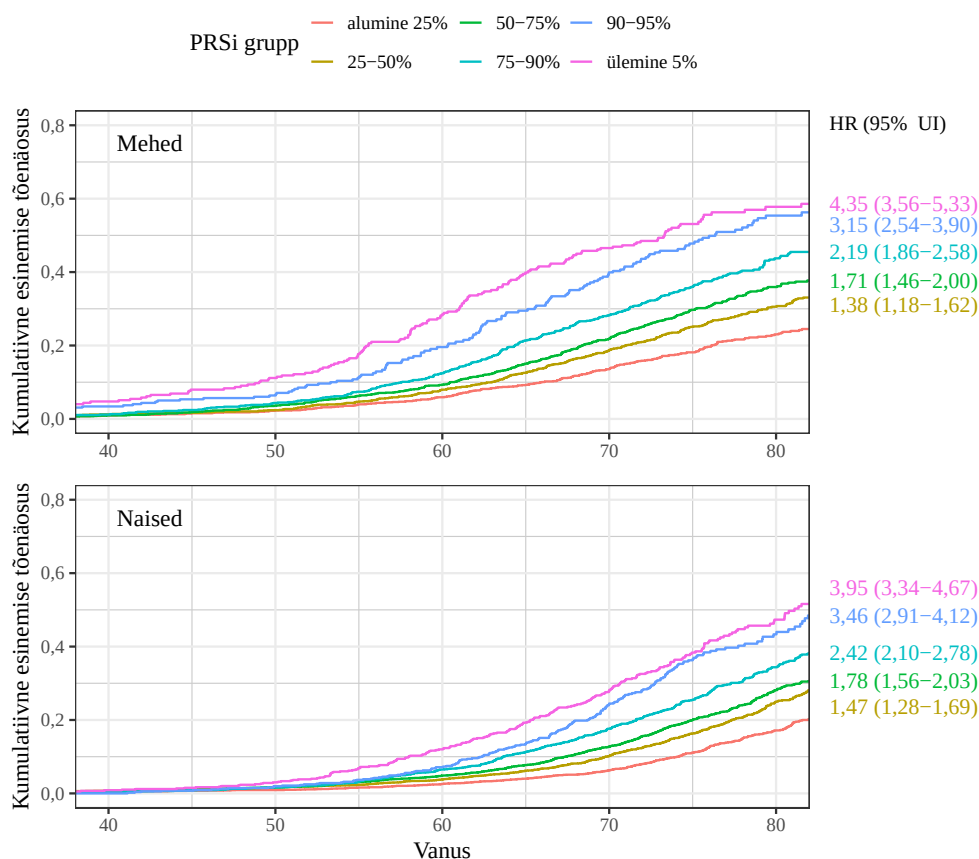
Joonis 2: Hinnangud KVA esinemusfunktsioonile soo ja kohordi lõikes

Mõlemas kohordis on meestel KVA hinnanguline kumulatiivne esinemise tõenäosus suurem kui naistel. Nii meeste kui naiste seas on esimesse kohorti (2002–2017 liitunud) kuuluvatel indiviididel hinnangulised kumulatiivsed esinemissagedused suuremad kui teise kohorti (2018–2021 liitunud) kuuluvatel indiviididel. Näiteks oli varasemas kohordis 65. eluaastaks kogenud KVAd hinnanguliselt ligikaudu 16,0% meestest (95% UI 15,0–17,0%), hilisemas kohordis aga umbes 10,5% (95% UI 9,7–11,4%). Naistel olid vastavad hinnangud 7,9% (95% UI 7,4–8,5%) ja 4,7% (95% UI 4,3–5,1%).

5.4.1 KVA risk PRSi protsentiilgruppides

Hindamaks polügeense riskiskoori mõju KVA riskile leiti Aalen-Johanseni hinnangud KVA esinemusfunktsioonile PRSi protsentiilide põhjal määratletud rühmades, eraldi soo ja kohordi lõikes. Vastavalt PRSi jaotusele jagati indiviidid kuude riskikategooriasse: esimesed kolm kvartiili defineerisid eraldi rühmad ja ülemine kvartiil jaotati omakorda kolmeks protsentiilivahemikuks (75%–90%, 90–95% ning ülemine 5%). Järgnevalt on iga soo- ja kohordipõhise grupi puhul kujutatud hinnangud KVA esinemusfunktsioonile, kus surma on käsitletud konkureeriva riskina, koos Coxi võrdeliste riskide mudeli põhjal hinnatud riskimäärade suhete ning vastavate 95% usaldusintervallidega, kus baastasemeks on alumine riskikategooria (esimene kvartiil).

Joonisel 3 on esitatud hinnangulist kumulatiivset esinemist kujutavad kõverad 2002.–2017. aastal liitunud indiviidide puhul. Kõigis geneetilise riskiskoori kategooriates on KVA risk statistiliselt oluliselt erinev alumisest kvartiilist ning kõrgema geneetilise riski korral suurem tõenäosus KVAsse haigestuda. Geneetilise riski ülemise 5% hulka kuuluvatel meestel hakkab tõenäosus saada KVA diagnoos enne teatud vanust teistest riskikategooriatest selgelt eristuma juba pärast 45. eluaastat. Naistel on aga kõigis PRSi riskikategooriates märgatavalt madalam KVA hinnanguline kumulatiivne esinemise tõenäosus kogu vaadeldud vanusevahemikus. Ülemise 5% hulka kuuluvate naiste hinnanguline KVA kumulatiivne esinemise tõenäosus eristub teistest riskirühmadest märgatavalt hilisemas eas kui meestel, ligikaudu 55. eluaastast.

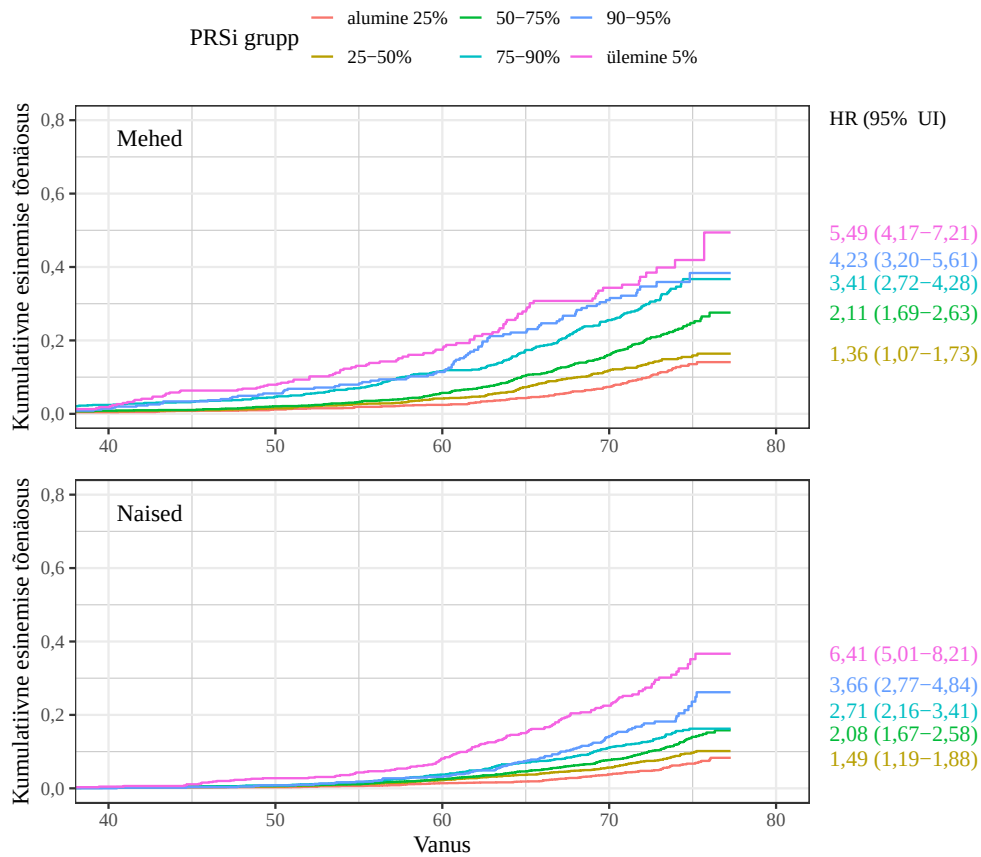


Joonis 3: Kohort 1 – hinnangud KVA esinemisfunktsioonile ja riskimäärade suhted PRSi protsentiiligruppide lõikes

Riskimäärade suhete põhjal on ülemise 5% hulka kuuluvatel meestel hinnanguliselt 4,35 korda suurem risk haigestuda KVAse võrreldes meestega alumises kvartiilis, samas kui ülemise 5% hulka kuuluvatel naistel on hinnanguliselt 3,95 korda suurem risk haigestuda KVAse võrreldes naistega alumises kvartiilis.

Joonisel 4 on kujutatud kumulatiivse esinemise tõenäosuse hinnangud 2018.–2021. aastal liitunud indiviidide seas. Võrreldes esimese kohordiga jäävad hinnangulised kumulatiivsed esinemissagedused nii meeste kui naiste puhul madalamale tasemele. Näiteks kui esimeses kohordis on ülemise 5% hulka kuuluvatel meestel 65. eluaastaks hinnatud kumulatiivseks KVA esinemise tõenäosuseks 39,9% (95% UI 34,4–46,3%), siis teises kohordis on vastav näitaja umbes 27,9% (95% UI 22,7–34,2%). Samas on PRSi kategooriate riskimäärade suhted alumise riskikategooriaga võrreldes märgatavalt suuremad kui esimeses kohordis. Kõrgeima riskiga rühmas (ülemine 5%)

olevatel meestel on haigestumise risk hinnanguliselt 5,49 korda suurem võrreldes madalaima riskiga rühmaga. Kõrgeima riskiga naistel on hinnanguliselt 6,41-kordne risk haigestuda KVAse võrreldes alumise kvartiiliga.



Joonis 4: Kohort 2 – hinnangud KVA esinemisfunktsioonile ja riskimäärade suhted PRSi protsentiiligruppide lõikes

Sarnaselt esimesele kohordile on ka teises kohordis naistel kõigis PRSi kategooriates madalam kumulatiivne haigestumistõenäosus kui meestel. Siiski on PRSi äärmuslike rühmade (alumine kvartiil vs ülemine 5%) vaheline suhteline riskide erinevus suurem naistel. Näiteks on ülemise 5% PRSi rühma kuuluvate meeste kumulatiivne haigestumistõenäosus 65. eluaastaks ligi 6,5 korda kõrgem kui madalaima riskiga rühmas (27,9% vs 4,3%), samas kui naistel on vastav erinevus pea kaheksakordne (15,0% vs 1,9%).

5.4.2 Coxi võrdeliste riskide mudelid

KVA esinemusfunktsiooni hinnangute põhjal (joonis 2) erineb KVA kumulatiivne esinemise tõenäosus nii soo kui kohordi lõikes ning erinevused muutuvad vanuse kasvades selgemaks. Et hinnata nii mittegeneetiliste riskifaktorite kui geneetilise riskiskoori mõju KVA tekkeriskile, võttes arvesse erinevusi kohortide, sugude ja vanuse lõikes, sobitati nende tunnuste alusel moodustatud gruppidele eraldi Coxi võrdeliste riskide mudelid. Mudelites kasutati ajaskaalana vanust ning arvestati paremalt tsenseerituse ja vasakult tõkestatusega. Kokku hinnati kuus mudelit järgmistes alamrühmades, kus vanus tähistab vanust geenivaramuga liitumisel:

- kohort 1, mehed, 18–59 aastat,
- kohort 2, mehed, 18–59 aastat,
- kohort 1, naised, 18–59 aastat,
- kohort 2, naised, 18–59 aastat,
- kohort 1, 60+ aastat (soo järgi kihitatud),
- kohort 2, 60+ aastat (soo järgi kihitatud).

Kuna vanemaealisi (60+ vanuses liitunuid) oli valimis vähem, siis ei hinnatud selles vanuserühmas soopõhiselt eraldi mudeleid, vaid kasutati soo järgi kihitamist, st meestele ja naistele lubati erinev baasrisk. Mudelid koostati paarikaupa sama soo ja vanuserühma lõikes. Iga sellise paari puhul sisaldasid lõplikud mudelid samu riskifaktoreid, erinevate soo ja vanuserühma kombinatsioonide vahel võisid lõplikesse mudelitesse jäävad tunnused erineda. Mudelite koostamist alustati täismudelist, mis sisaldas kõiki käsitletavaid riskitegureid. Lõplikku mudelisse jäeti tunnused, mis olid vähemalt ühes vastava paari mudelis statistiliselt olulised.

Tabelis 4 on esitatud Coxi mudeli parameetrite hinnangud 18–59-aastaselt liitunud meeste seas kahes kohordis, koos standardvigade, riskimäärade suhete ning neile vastavate 95% usaldusintervallidega. Suitsetamise puhul on baastasemeks mitte-suitsetamine ning haiguste puhul diagnoosi puudumine.

Tabel 4: KVA riskitegurid 18–59-aastaselt liitunud meestel

	$\hat{\beta}$ (se)	HR (95% UI)
Kohort 1, mehed, 18–59		
PRS	0,447 (0,033)	1,56 (1,47–1,67)
Suitsetamine	0,249 (0,065)	1,28 (1,13–1,46)
KMI	0,045 (0,007)	1,05 (1,03–1,06)
Südamepuudulikkus	0,561 (0,107)	1,75 (1,42–2,16)
Kõrgvererõhkhaigused	0,376 (0,072)	1,46 (1,26–1,68)
Kohort 2, mehed, 18–59		
PRS	0,630 (0,046)	1,88 (1,71–2,06)
Suitsetamine*	-0,021 (0,111)	0,98 (0,79–1,22)
KMI	0,063 (0,009)	1,06 (1,05–1,08)
Südamepuudulikkus	0,945 (0,186)	2,57 (1,79–3,71)
Kõrgvererõhkhaigused	0,463 (0,106)	1,59 (1,29–1,96)

* statistiliselt ebaoluline tunnus ($p > 0,05$)

18–59-aastaselt liitunud meeste puhul osutusid statistiliselt olulisteks KVA riski mõjutavateks tunnusteks PRS, KMI ning südamepuudulikkuse ja kõrgvererõhkhaiguste esinemine, esimeses kohordis ka suitsetamine. Kõigi mõlemas kohordis oluliste tegurite mõju suunad olid kohortide vahel kooskõlas – kõik tegurid suurendavad KVA riski –, kuid efektide tugevused olid suuremad teises kohordis (2018–2021 liitunud). Oluline erinevus kahe kohordi vahel ilmnis suitsetamise puhul. Kui esimeses kohordis oli suitsetamine statistiliselt oluline ning seotud hinnanguliselt 1,28 korda kõrgema riskiga võrreldes mitesuitsetajatega, siis teises kohordis ei olnud seos suitsetamise ja KVA esinemise vahel statistiliselt oluline. Enim suurendasid KVA riski mõlemas kohordis südamepuudulikkuse esinemine ning geneetiline risk. Kui võrrelda kahte indiviidi, kelle PRS erineb ühe standardhälbe võrra, siis on kõrgema geneetilise riskiga indiviidil esimeses kohordis hinnanguliselt 1,56 korda ning teises kohordis 1,88 korda suurem KVAse haigestumise risk. Sarnane muster ilmnis ka südamepuudulikkuse korral: kui esimeses kohordis oli haiguse olemasolu seotud hinnanguliselt 1,75 korda suurema KVA riskiga võrreldes diagnoosita indiviididega, siis teises kohordis oli vastav hinnang oluliselt kõrgem, ligikaudu 2,57.

Tabelis 5 on esitatud 18–59-aastaselt liitunud naiste KVA riski mõjutavad tegurid kahes kohordis. Hariduse puhul on baastasemeks alla keskhariduse omamine.

Tabel 5: KVA riskitegurid 18–59-aastaselt liitunud naistel

	$\hat{\beta}$ (se)	HR (95% UI)
Kohort 1, naised, 18–59		
PRS	0,490 (0,030)	1,63 (1,54–1,73)
Haridus		
Keskharidus*	-0,139 (0,107)	0,87 (0,71–1,07)
Kõrgharidus	-0,269 (0,116)	0,76 (0,61–0,96)
KMI	0,045 (0,005)	1,05 (1,04–1,06)
Südamepuudulikkus	0,576 (0,108)	1,78 (1,44–2,20)
Kõrgvererõhkhaigused	0,441 (0,069)	1,55 (1,36–1,78)
Kohort 2, naised, 18–59		
PRS	0,596 (0,049)	1,81 (1,65–2,00)
Haridus		
Keskharidus*	-0,343 (0,235)	0,71 (0,45–1,13)
Kõrgharidus	-0,749 (0,240)	0,47 (0,30–0,76)
KMI	0,035 (0,009)	1,04 (1,02–1,05)
Südamepuudulikkus	0,914 (0,192)	2,50 (1,71–3,63)
Kõrgvererõhkhaigused	0,555 (0,115)	1,74 (1,39–2,18)

* statistiliselt ebaoluline tunnus ($p > 0,05$)

18–59-aastaselt liitunud naiste puhul olid mõlemas kohordis KVA riski mõjutavad tegurid PRS, haridustase, KMI ning südamepuudulikkuse ja kõrgvererõhkhaiguste esinemine. Sarnaselt samas vanuses meestele, olid kõigi tegurite mõju suunad kohortide vahel kooskõlas ning efektide tugevused üldjuhul suuremad teises kohordis. Nii PRS, KMI kui ka südamepuudulikkuse ja kõrgvererõhkhaiguste esinemine olid seotud suurenenud KVA riskiga. Erinevalt meestest ei osutunud suitsetamine naiste mudelites statistiliselt oluliseks, samas kui oluline kaitsva mõjuga tegur oli haridustase – kõrghariduse omamine oli mõlemas kohordis seotud madalama KVA riskiga võrreldes alla keskhariduse omamisega. Kõrgharidusega indiviidide KVA risk oli esimeses kohordis hinnanguliselt 1,31 korda väiksem ning teises kohordis lausa 2,12

korda väiksem võrreldes alla keskharidusega indiviididega, samas kui keskharidusega ja alla keskharidusega indiviidide riskitase ei erinenud kummaski kohordis. Sarnaselt meeste mudelitele avaldasid KVAse haigestumise riskile suurimat mõju geneetiline risk ja südamepuudulikkuse esinemine.

Vähemalt 60-aastaselt liitunud indiviidide KVA riski mõjutavad tegurid kohorditi on esitatud tabelis 6.

Tabel 6: KVA riskitegurid vähemalt 60-aastaselt liitunudatel

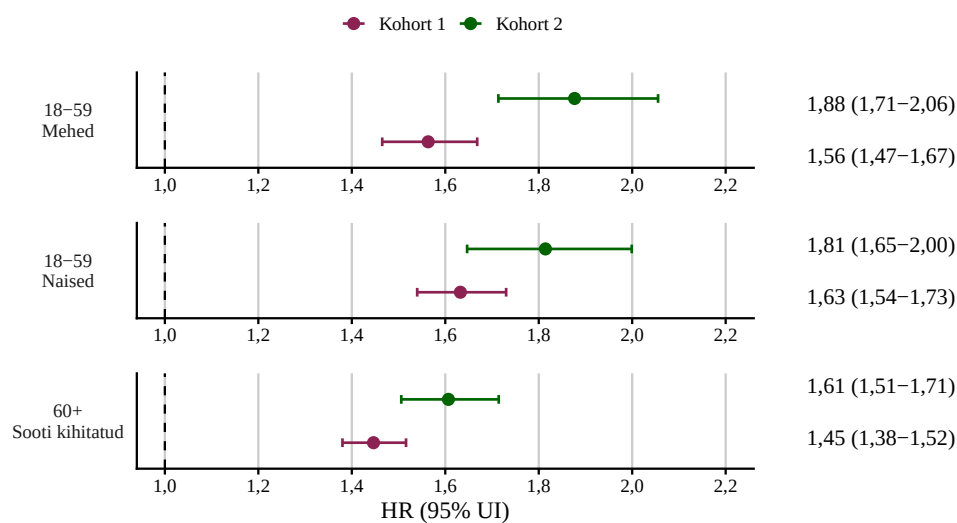
	$\hat{\beta}$ (se)	HR (95% UI)
Kohort 1, 60+		
PRS	0,369 (0,024)	1,45 (1,38–1,52)
Haridus		
Keskharidus*	-0,057 (0,059)	0,94 (0,84–1,06)
Kõrgharidus	-0,202 (0,070)	0,82 (0,71–0,94)
KMI	0,026 (0,005)	1,03 (1,02–1,04)
I tüüpi diabeet	0,345 (0,125)	1,41 (1,10–1,80)
Südamepuudulikkus	0,370 (0,056)	1,45 (1,30–1,62)
Orgaanilised unehäired*	0,064 (0,076)	1,07 (0,92–1,24)
Kõrgvererõhkhaigused	0,249 (0,057)	1,28 (1,15–1,43)
Kohort 2, 60+		
PRS	0,474 (0,033)	1,61 (1,51–1,71)
Haridus*		
Keskharidus	-0,220 (0,121)	0,80 (0,63–1,02)
Kõrgharidus	-0,219 (0,124)	0,80 (0,63–1,03)
KMI	0,033 (0,006)	1,03 (1,02–1,05)
I tüüpi diabeet*	0,112 (0,210)	1,12 (0,74–1,69)
Südamepuudulikkus	0,475 (0,093)	1,61 (1,34–1,93)
Orgaanilised unehäired	0,275 (0,071)	1,32 (1,14–1,51)
Kõrgvererõhkhaigused	0,272 (0,075)	1,31 (1,13–1,52)

* statistiliselt ebaoluline tunnus ($p > 0,05$)

Vähemalt 60-aastaste seas olid mõlemas kohordis KVA riski suurendavad tegurid PRS, KMI ning südamepuudulikkuse ja kõrgvererõhkhaiguste esinemine. Lisaks osu-

tusid sõltuvalt kohordist oluliseks ka haridustase ning I tüüpi diabeedi ja orgaaniliste unehäirete esinemine. Sarnaselt noorematega olid kõigi mõlemas kohordis oluliste tegurite mõju suunad kooskõlas ning efektide tugevused suuremad teises kohordis. Võrreldes noorema vanuserühma mudelitega ilmnes vähemalt 60-aastaste seas suurem hulk haigusi, millel oli mõju KVA avaldumisele – kui noormate seas suurendasid haigustest oluliselt KVA riski vaid südamepuudulikkus ja kõrgvererõhkhaigused, siis vanemate seas lisandusid kohorditi ka I tüüpi diabeet ja orgaanilised unehäired. Esimeses kohordis oli I tüüpi diabeedi esinemine seotud hinnanguliselt 1,41 korda suurema KVA riskiga võrreldes diagnoosita indiviididega ning kõrghariduse olemisel oli kaitsev mõju. 2018.–2021. aastal liitunud seas ei osutunud I tüüpi diabeedi diagnoosi olemasolu ega haridustase olulisteks KVA riski mõjutavateks teguriteks, samas kui orgaaniliste unehäirete esinemine oli seotud hinnanguliselt 1,32 korda suurema KVA riskiga võrreldes diagnoosita indiviididega. Sarnaselt noormate mudelitele avaldasid vähemalt 60-aastaste seas KVA haigestumise riskile suurimat mõju geneetiline risk ja südamepuudulikkuse esinemine.

Joonisel 5 on kokkuvõtlikult kujutatud PRSi riskimäärade suhted Coxi mudelites koos 95% usaldusintervallidega. PRSi mõju on suurem noormate ning teise kohordi puhul, seejuures 18–59-aastaste meeste puhul on PRSi mõju teises kohordis oluliselt erinev esimese kohordi omast (usaldusintervallid ei kattu).



Joonis 5: PRSi riskimäärade suhted Coxi mudelites

5.4.3 Mudelite võrdeliste riskide eelduse kontrollimine

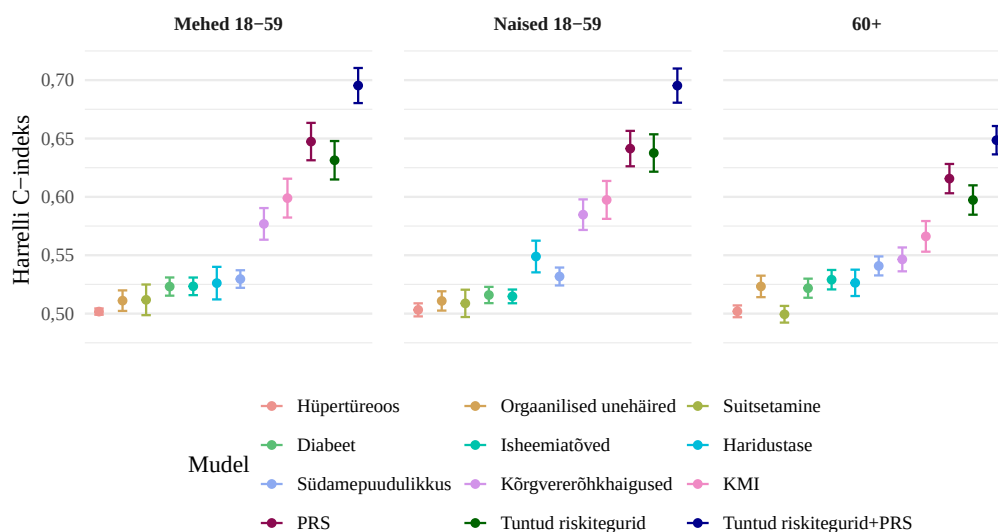
Coxi mudelite võrdeliste riskide eeldust kontrolliti kaalutud Schoenfeldi jääkide abil. Lisaks statistilisele olulisusele vaadeldi ka jääkide graafikuid, et hinnata kovariaatide mõjude võimalikke muutusi ajas. Tabel kaalutud Schoenfeldi jääkide testi tulemustega ning jääkide graafikute näited on esitatud lisa 2.

Kuuest mudelist nelja puhul viitas kaalutud Schoenfeldi jääkidel põhinev test mõningatele kõrvalekalletele võrdeliste riskide eeldusest. Statistiliselt olulised kõrvalekalded ilmneshid peamiselt PRSi, südamepuudulikkuse ja kõrgvererõhkaiguste puhul. Jääkide graafikute visuaalne hindamine näitas siiski, et kõrvalekalded olid üldiselt väikesed ning väljendusid ajas sujuvate ja nõrkade trendidena, viidates enamasti mõjude kergelt vähenevale suundumusele vanuse kasvades, mis on kooskõlas vanema vanuserühma mudelites täheldatud madalamate efektihiinangutega. Arvestades, et täheldatud kõrvalekalded olid väikesed ega viidanud järskudele või mittelineaarsetele muutustele ajas, võib võrdeliste riskide eeldust pidada analüüsi kontekstis piisavalt täidetuks ja tõlgendada Coxi mudelite riskimäärade suhteid kui keskmisi mõjusid jälgimisperioodi vältel.

5.4.4 C-indeksite võrdlus

Et hinnata, kui täpselt eristavad traditsioonilised (mittegeneetilised) riskitegurid ja PRS üksikult ning kombineeritult KVA tekkeriski, võrreldi eri mudelite diskrimineerimisvõimeid Harrelli C-indeksi alusel. Selleks sobitati Coxi võrdeliste riskide mudelid vanuse- ja soopõhiste rühmade kaupa, mis olid kohordi järgi kihitatud (60+ vanuserühmas täiendavalt ka soo järgi). Tulemusi, kus mudelid sobitati iga kohordi jaoks eraldi, pole esitatud, kuna täiendav analüüs näitas, et peamised järeldused mudelite diskrimineerimisvõimete kohta jäid samaks ka vanuse-, soo- ja kohordipõhiste rühmade käsitlemise korral. Igas grupis hinnati kokku 12 mudelit: ühe argumenttunnusega mudelid iga traditsioonilise riskiteguri ja PRSi jaoks eraldi, mitmene mudel, mis sisaldas kõiki traditsioonilisi riskitegureid, ning mudel, mis hõlmas nii kõiki tuntud riskitegureid kui ka PRSi. Traditsiooniliste riskitegurite puhul liideti diabeedi diagnoosid (E10 ja E11), samuti südame isheemiatõved (I21, I22 ja

I25). Mudelitele vastavad C-indeksid koos 95% usaldusintervallidega on kujutatud joonisel 6, täpsed C-indeksite väärtused on esitatud lisas 3.



Joonis 6: Harrelli C-indeksite väärtused

Kõigis gruppides jäid traditsiooniliste riskitegurite ühekovariaadiliste mudelite Harrelli C-indeksite väärtused alla 0,6, mis viitab piiratud diskrimineerimisvõimele. Seevastu PRS osutus kõigis gruppides tugevaimaks üksikuks ennustajaks ja sellel põhineva mudeli diskrimineerimisvõime oli samaväärne või parem kui mudelil, mis sisaldas kõiki tuntud riskitegureid. Parim diskrimineerimisvõime saavutati kõigis gruppides mudeli puhul, mis kombineeris traditsioonilised riskitegurid ja PRSi.

6 Arutelu

Analüüsist selgus, et kodade virvendusarütmia ja laperduse tekkerisk on oluliselt seotud nii mitmete mittegeneetiliste riskitegurite kui ka polügeense riskiskooriga. Järgnevalt arutletakse peamiste tulemuste üle, keskendudes esmalt mittegeneetilistele riskiteguritele, seejärel PRSi rollile ja lõpuks analüüsi piirangutele.

KVA tekkerisk on tugevalt seotud kehamassiindeksi ning kõrgvererõhkhaiguste ja südamepuudulikkuse esinemisega, mida käsitletakse ka kliinilises praktikas oluliste riskiteguritena. Nende tegurite olulisus kõigis kohordi-, soo- ja vanuserühmapõhistes gruppides viitab nende kesksele rollile KVA avaldumises. Samas ei olnud mõnede tuntud riskitegurite, nagu suitsetamise, I tüüpi diabeedi ja orgaaniliste unehäirete mõju kõigis gruppides statistiliselt oluline. Ka kõrgema haridustaseme kaitsev efekt ilmnis vaid osades gruppides. Need erinevused KVA riski mõjutavates tunnustes võivad ühelt poolt tuleneda piiratud statistilisest võimsusest, näiteks esines I tüüpi diabeedi diagnoosiga indiviide mõnes alarühmas vähe. Teisalt ei saa välistada, et nimetatud tegurite mõju varieerub soo ja vanuse lõikes või on osaliselt kirjeldatud teiste mudelitesse kaasatud riskitegurite kaudu. Riskitegurite efektide suurused olid üldjuhul suuremad teises kohordis (2018–2021 liitunud) ning nooremas vanuserühmas. Kohortidevahelised erinevused võivad osaliselt peegeldada värbamismeetodite ja seega kohortide koosseisude erinevusi, aga ka asjaolu, et kohortidesse kuuluvad eri generatsioonide esindajad. Vanuserühmade vaheline muster võib tuleneda sellest, et vanemaealiste puhul ilmnis rohkem kaasuvaid haigusi, millel oli mõju KVA avaldumisele. See võib viidata sellele, et nooremas eas avalduv KVA on tugevamini seotud üksikute domineerivate riskiteguritega, samas kui eakamatel esineb sagedamini samaaegselt mitmeid kaasuvaid haigusi, mis võib üksikute tegurite mõju vähendada.

PRSi analüüs näitas, et geneetiline eelsoodumus on KVA tekkeriski oluline ennustaja. Kõigis soo- ja kohordipõhistes gruppides ilmnis, et kõrgema geneetilise riskiga isikutel on suurem tõenäosus KVAse haigestuda võrreldes madalaima riskikategooriaga. Seejuures kõrgeimasse geneetilise riski kategooriasse (ülemine 5%) kuuluvatel indiviididel oli KVAse haigestumise risk sõltuvalt kohordist ja soost

hinnanguliselt 3,95 kuni 6,41 korda suurem võrreldes madalaima riskiga indiviididega. Tegemist on märkimisväärse erinevusega, mis viitab PRSi potentsiaalsele kasulikkusele just personaalse ennetuse kontekstis. Oluline on ka asjaolu, et PRS jäi kõigis vaadeldavates gruppides oluliseks ka pärast mittegeneetiliste tegurite arvesse võtmist, mis viitab sellele, et geneetiline risk kannab lisainformatsiooni, mida traditsioonilised riskitegurid ei kata. PRS osutus ka tugevaimaks üksikuks KVA tekke ennustajaks, olles diskrimineerimisvõimelt samaväärne või parem kui kõigi mittegeneetiliste tegurite kombinatsioon. Samas saavutati parim diskrimineerimisvõime traditsiooniliste riskitegurite ja PRSi kombineerimisel. Seega on nii geneetilise eelsoodumuse kui mittegeneetiliste tegurite integreerimine oluline terviklikuma riskihindamise saavutamiseks – PRS ei asenda kliinilisi riskitegureid, vaid täiendab neid. Selline kombineeritud lähenemine võib olla eriti väärtuslik varajase ennetuse kontekstis, kus geneetiline risk võib aidata tuvastada kõrge riskiga indiviide enne kliiniliste riskitegurite avaldumist.

Tulemuste tõlgendamisel tuleb arvestada mitmete piirangutega. Võrdeliste riskide eeldus ei olnud Coxi mudelites mõne tunnuse puhul täidetud, mis tähendab, et teatud riskitegurite mõju KVA tekkeriskile võib ajas muutuda. Seetõttu võiks tulevastes uuringutes kaaluda ajast sõltuvate kovariaatide lisamist mudelitesse või eraldi mudelite loomist enamatele alamrühmadele. Lisaks ei arvestatud Coxi mudelites surma kui konkureeriva riskiga, erinevalt Aalen-Johanseni hinnangutest. Kuigi on näidatud, et ka sel juhul on võimalik mudeli parameetrid korrektselt hinnata ning tõlgendada riskimäärade suhteid põhjusespetsiifilistena (Therneau, Crowson, Atkinson, 2026), oleks soovituslik rakendada konkureerivaid riske arvestavaid mudeleid, näiteks Fine ja Gray võrdeliste alamriskide meetodit.

Kokkuvõte

Bakalaureusetöö eesmärk oli tuvastada kodade virvendusarütmia ja laperduse tekke-riskiga seotud mittegeneetilised riskifaktorid ning eelkõige hinnata geneetilise riskiskoori lisandväärtust prognoosimudelites. Analüüsi läbiviimiseks kasutati Tartu Ülikooli Eesti Geenivaramu andmeid. KVA potentsiaalsete riskifaktoritena kaasati töös lisaks geneetilisele riskiskoorile kõrgeim omandatud haridustase, suitsetamise staatus, kehamassiindeks ning mitmed kaasuvad haigused, sealhulgas kilpnäärme haigusseisundid, diabeet, kõrgvererõhkhaigused, südame isheemiatõved ja unehäired.

Coxi võrdeliste riskide mudelitest, mis hinnati soo-, vanuse- ja kohordipõhistes alamrühmades, ilmnes, et kehamassiindeks ning kõrgvererõhkhaiguste ja südamepuudulikkuse esinemine olid järjepidevalt seotud suurema KVA tekkeriskiga. Samas ei olnud kõigi tunnuste seos KVA tekkeriskiga ühtlaselt oluline – suitsetamise, haridustaseme, I tüüpi diabeedi ning orgaaniliste unehäirete mõju ilmnes vaid osades gruppides. Samuti varieerusid kohortide ja vanuserühmade lõikes riskitegurite mõjude suurusjärgud, olles üldjuhul suuremad hilisemas kohordis ning nooremate seas.

Geneetilise riskiskoori lisamine prognoosimudelitesse andis selget lisaväärtust võrreldes üksnes mittegeneetiliste tegurite käsitlemisega. Aalen-Johanseni hinnangutest ja Coxi võrdeliste riskide mudelite põhjal hinnatud riskimäärade suhetest selgus, et kõrgema geneetilise riskiga indiviididel oli märkimisväärselt suurem haigestumiskrisk võrreldes madalaima riskikategooriaga. Seejuures kinnitas PRSi iseseisvat panust riskihindamisse ka asjaolu, et PRS jäi oluliseks mudelites, kuhu olid kaasatud mittegeneetilised riskitegurid. Võrreldes Harrelli C-indeksi põhjal, kuivõrd mittegeneetilised riskitegurid ja PRS üksikult ning kombineeritult KVA tekkeriski eristavad, nähtus, et PRS eraldi oli vähemalt sama hea või parem kui kõigi mittegeneetiliste tegurite kombinatsioon. Parim diskrimineerimisvõime saavutati siiski mudelites, mis hõlmasid nii mittegeneetilisi riskitegureid kui PRSi. Tulemuste tõlgendamisel tuleb aga arvestada, et võrdeliste riskide eeldus ei olnud kõigis Coxi mudelites täidetud ning mudelites ei käsitletud surma konkureeriva riskina.

Kodade virvendusarütmia ja laperduse riskihindamine on seega terviklikum, kui

kombineerida mittegeneetilised riskitegurid geneetilise riskiga – PRS kannab lisainformatsiooni, mida mittegeneetilised tegurid ei pruugi katta ning võib võimaldada kõrge riskiga indiviidide tuvastamist ka enne kliiniliste riskifaktorite avaldumist. See viitab kombineeritud lähenemise väärtuslikkusele just varajase avastamise ja personaalse ennetuse kontekstis.

Kasutatud allikad

Alonso, A. *et al.* (2013). „Simple Risk Model Predicts Incidence of Atrial Fibrillation in a Racially and Geographically Diverse Population: the CHARGE-AF Consortium“. *JAHA*. 2013.2, lk 1–11. DOI: 10.1161/JAHA.112.000102.

American Heart Association. (2025). *Atrial Flutter*. URL: <https://www.heart.org/en/health-topics/atrial-fibrillation/what-is-atrial-fibrillation-a-fib-or-af/atrial-flutter> (vaadatud 28.01.2026).

Andersen, J. H., L. Andreassen, M. S. Olesen. (2020). „Atrial fibrillation – a complex polygenic disease“. *European Journal of Human Genetics*. 2020.29, lk 1051–1060. DOI: 10.1038/s41431-020-00784-8.

Choi, S. W., T. S.-H. Mak, P. F. O'Reilly. (2020). „Tutorial: a guide to performing polygenic risk score analyses“. *Nature Protocols*. 2020.15, lk 2759–2772. DOI: 10.1038/s41596-020-0353-1.

Collett, D. (2014). „Modelling Survival Data in Medical Research (3rd ed.)“. Chapman and Hall/CRC. DOI: 10.1201/b18041.

Frazer, K. A., S. S. Murray, N. J. Schork, E. J. Topol. (2009) „Human genetic variation and its contribution to complex traits“. *Nature Reviews Genetics*. 2009.10, lk 241–251. DOI: 10.1038/nrg2554.

Gunn, S. *et al.* (2024). „Comparison of methods for building polygenic scores for diverse populations“. *Human Genetics and Genomics Advances*. 2024.6, lk 1–11. DOI: 10.1016/j.xhgg.2024.100355.

Harrell, F. E., K. L. Lee, D. B. Mark. (1996). „Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors“. *Statistics in Medicine*. 1996.15, lk 361–387. DOI: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

Hosmer, D. W., S. Lemeshow, R. X. Sturdivant. (2013). „Applied Logistic Regression (3rd ed.)“. Wiley. DOI: 10.1002/9781118548387.

Järving, K. L., T. Teeäär, M. Elmet, E. Laane, U. Arus. (2021). *Kodade virvendusarütmia*. Tartu Ülikooli Kliinikum. URL: <https://www.kliinikum.ee/patsiendi>

info-andmebaas/kodade-virvendusarutmia/ (vaadatud 28.01.2026).

Kleinbaum, D. G., M. Klein. (1997). „Survival Analysis: A Self-Learning Text“. Springer. DOI: 10.1007/978-1-4419-6646-9.

Läll, K. (2019). *Risk scores and their predictive ability for common complex diseases*. Doktoritöö. Tartu Ülikool. URL: <http://hdl.handle.net/10062/63766> (vaadatud 22.11.2025).

Migdady, I., A. Russman, A. B. Buletko. (2021). „Atrial Fibrillation and Ischemic Stroke: A Clinical Review“. *Seminars in Neurology*. 2021.41, lk 348–364. DOI: 10.1055/s-0041-1726332.

Milani, L. *et al.* (2025). „The Estonian Biobank’s journey from biobanking to personalized medicine“. *Nature Communications*. 2025.16, lk 1–13. DOI: 10.1038/s41467-025-58465-3.

Moorthie, S. *et al.* (2021). *Polygenic scores and clinical utility*. PHG Foundation. URL: <https://www.phgfoundation.org/resources/reports/polygenic-scores-and-clinical-utility/> (vaadatud 10.11.2025).

National Cancer Institute. (*s. a.*). *Single nucleotide polymorphism (SNP)*. URL: <https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/snp> (vaadatud 10.11.2025).

National Human Genome Research Institute. (2020). *Polygenic risk scores*. URL: <https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores> (vaadatud 10.11.2025).

Nielsen, J. B. *et al.* (2018). „Genome-wide Study of Atrial Fibrillation Identifies Seven Risk Loci and Highlights Biological Pathways and Regulatory Elements Involved in Cardiac Development“. *The American Journal of Human Genetics*. 2018.102, lk 103–115. DOI: 10.1016/j.ajhg.2017.12.003.

Polygenic Risk Score Task Force of the International Common Disease Alliance. (2021). „Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps“. *Nature Medicine*. 2021.27, lk 1876–1884. DOI: 10.1038/s41591-021-01549-6.

Polygenic Score (PGS) Catalog. (s. a.). URL: <https://www.pgscatalog.org> (vaadatud 15.04.2026).

Ravijuhend. (2021). *Kodade virvendusarütmia (KVA) kui haiguse olemus, vormid, tüsistused, sümptomid, diagnoosimine*. URL: <https://www.ravijuhend.ee/patsiendivarav/juhendid/183/kodade-virvendusarutmia-kva-kui-haiguse-olemus-vormid-tusistused-sumptomid-diagnoosimine> (vaadatud 27.01.2026).

Regionaalhaigla Südamekeskus. (s. a.). *Rütmihäired*. URL: <https://sudamekeskus.ee/valdkond/rutmihaired> (vaadatud 28.01.2026).

Roselli, C., M. Rienstra, P. T. Ellinor. (2020). „Genetics of Atrial Fibrillation in 2020: GWAS, Genome Sequencing, Polygenic Risk, and Beyond“. *Circulation Research*. 2020.127, lk 21–33. DOI: 10.1161/CIRCRESAHA.120.316575.

Segan, L. *et al.* (2025). „Combining polygenic and clinical risk scores in atrial fibrillation risk prediction: Implications for population screening.“ *Heart Rhythm*. 2025.22, lk 1906–1914. DOI: 10.1016/j.hrthm.2025.04.032.

Therneau, T. M. (2026). *A package for survival analysis in R*. URL: <https://CRAN.R-project.org/package=survival> (vaadatud 27.02.2026).

Therneau, T. M., E. Atkinson. (2026). *Concordance*. URL: <https://cran.r-project.org/web/packages/survival/vignettes/concordance.pdf> (vaadatud 08.03.2026).

Therneau, T. M., C. Crowson, E. Atkinson. (2026). *Multi-state models and competing risks*. URL: <https://cran.r-project.org/web/packages/survival/vignettes/compete.pdf> (vaadatud 26.03.2026).

Thiébaud, A. C. M., J. Bénichou. (2004). „Choice of time-scale in Cox’s model analysis of epidemiologic cohort data: a simulation study.“ *Statistics in Medicine*. 2004.23, lk 3803–3820. DOI: 10.1002/sim.2098.

Torkamani, A., N. E. Wineinger, E. J. Topol. (2018). „The personal and clinical utility of polygenic risk scores“. *Nature Reviews Genetics*. 2018.19, lk 581–590. DOI: 10.1038/s41576-018-0018-x.

TÜ Genoomika Instituut. (s. a.). *Eesti geenivaramu*. URL: <https://genomics.ut.ee>

ee/et/eesti-geenivaramu (vaadatud 07.03.2026).

Van Gelder, I. C. *et al.* (2024). „2024 ESC Guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS)“. *European Heart Journal*. 2024.45, lk 3314–3414. DOI: 10.1093/eurheartj/ehae176.

Wray, N. R., M. E. Goddard, P. M. Visscher. (2007). „Prediction of individual genetic risk to disease from genome-wide association studies.“ *Genome Research*. 2007.17, lk 1520–1528. DOI: 10.1101/gr.6665407.

Lisad

Lisa 1. Elukestusanalüüsi meetodite rakendamine R abil

Elukestusandmete analüüsimiseks kasutatakse tarkvaras R paketti *survival*. Analüüsi aluseks on elukestusobjekt, mis kirjeldab iga indiviidi jälgimisaega ja seda, kas huvipakkuv sündmus leidis aset. Kui ajaskaalana kasutatakse vanust, luuakse funktsiooni *Surv* abil objekt kujul *Surv(vanus1, vanus2, sündmus)*, kus *vanus1* tähistab indiviidi vanust uuringusse sisenemisel ja *vanus2* vanust sündmuse toimumise või tsenseerimise hetkel. Tunnus *sündmus* näitab, kas huvipakkuv sündmus toimus (väärtus 1) või vaatlus oli tsenseeritud (väärtus 0). (Therneau, 2026)

Elulemusfunktsiooni Kaplan-Meieri hinnangu leidmiseks kasutatakse funktsiooni *survfit*, millele antakse ette valem kujul *Surv(vanus1, vanus2, sündmus) ~ grupp*. Vasakul poolel paikneb elukestust kirjeldav *Surv* objekt ja paremal rühmitav tunnus. Kui hinnangut soovitakse leida kogu valimi kohta, kasutatakse valemi paremal poolel konstanti 1. (Therneau, 2026)

Coxi võrdeliste riskide mudel sobitatakse funktsiooniga *coxph*, millele etteantav valem on kujul *Surv(vanus1, vanus2, sündmus) ~ x₁ + ... + x_p*. Valemi vasakul poolel paikneb elukestusobjekt ja paremal selgitavad tunnused, mille mõju elulemusele hinnatakse. Kihitatud mudeli korral lisatakse valemi paremale poolele ka argument *strata(tunnus)*, kus *tunnus* määrab kihid. Mudeli võrdeliste riskide eelduse kontrollimiseks kaalutud Schoenfeldi jääkide abil kasutatakse funktsiooni *cox.zph*, mille argument on funktsiooni *coxph* abil sobitatud Coxi mudel. Mudeli C-indeksi arvutamiseks rakendatakse sobitatud Coxi mudelile funktsiooni *concordance*. (Therneau, 2026)

Aalen-Johanseni hinnangu leidmiseks kasutatakse sarnaselt Kaplan-Meieri hinnangule funktsiooni *survfit*, kuid elukestusobjekti kuju erineb varem defineeritust. Surma kui konkureerivat riski arvesse võttes luuakse funktsiooni *Surv* abil objekt kujul *Surv(vanus1, vanus2, seisund)*, kus tunnusel *seisund* on kolm taset: väärtus 0 tähistab tsenseeritud vaatlust, väärtus 1 huvipakkuva sündmuse toimumist ja väärtus 2 konkureeriva sündmuse (surma) toimumist. Funktsioonile *survfit* etteantav valem

on seega kujul $Surv(vanus1, vanus2, seisund) \sim grupp$. (Therneau, 2026)

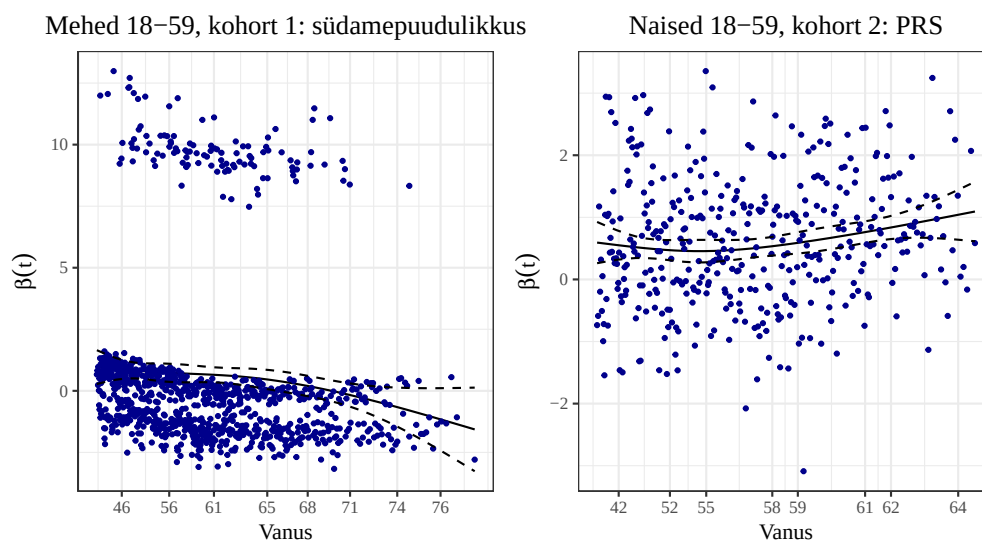
Lisa 2. Coxi võrdeliste riskide mudelite eelduse kontrollimine

Tabel L1: Kaalutud Schoenfeldi jääkide testi p-väärtused

Tunnus	Mehed 18–59		Naised 18–59		60+	
	Kohort 1	Kohort 2	Kohort 1	Kohort 2	Kohort 1	Kohort 2
PRS	0,0048*	0,2653	0,2613	0,0140*	0,0052*	0,1109
Suitsetamine	0,6781	0,4201	–	–	–	–
Haridus	–	–	0,8646	0,0995	0,4350	0,6517
KMI	0,0976	0,5348	0,4293	0,6512	0,0192*	0,5105
I tüüpi diabeet	–	–	–	–	0,8145	0,8267
Südamepuudulikkus	$2,03 \times 10^{-5}$ *	0,1889	0,0003*	0,0218*	0,0061*	0,0845
Orgaanilised unehäired	–	–	–	–	0,2241	0,6314
Kõrgvererõhahaigused	0,0287*	0,8537	0,0196*	0,0974	0,5303	0,7200

* statistiliselt oluline kõrvalekalle võrdeliste riskide eeldusest ($p < 0,05$)

Joonisel L1 on näitena esitatud esimese kohordi nooremate meeste mudelis südamepuudulikkuse ja teise kohordi nooremate naiste mudelis PRSi kaalutud Schoenfeldi jääkide graafikud, mis illustreerivad nende kovariaatide mõju võimalikku muutumist ajas. Esimene neist on esitatud, kuna südamepuudulikkus rikkus enim võrdeliste riskide eeldust, ja teine, kuna PRSi puhul täheldati kasvavat trendi.



Joonis L1: Kaalutud Schoenfeldi jääkide graafikute näited

Lisa 3. Harrelli C-indeksite väärtused

Tabel L2: Harrelli C-indeksite väärtused ja 95% usaldusintervallid

Mudel	Mehed 18–59	Naised 18–59	60+
Hüpertürees	0,502 (0,499–0,504)	0,503 (0,498–0,509)	0,502 (0,497–0,507)
Organilised unehäired	0,511 (0,502–0,520)	0,511 (0,503–0,519)	0,523 (0,514–0,532)
Suitsetamine	0,512 (0,499–0,525)	0,509 (0,497–0,520)	0,499 (0,492–0,507)
Diabeet	0,523 (0,515–0,531)	0,516 (0,509–0,523)	0,522 (0,514–0,530)
Isheemiatõved	0,523 (0,516–0,531)	0,515 (0,509–0,521)	0,529 (0,521–0,537)
Haridustase	0,526 (0,512–0,540)	0,549 (0,535–0,562)	0,526 (0,515–0,538)
Südamepuudulikkus	0,530 (0,522–0,537)	0,532 (0,524–0,539)	0,541 (0,533–0,549)
Kõrgvererõhkhaigused	0,577 (0,563–0,590)	0,585 (0,572–0,598)	0,546 (0,536–0,557)
KMI	0,599 (0,582–0,616)	0,597 (0,581–0,614)	0,566 (0,553–0,579)
PRS	0,647 (0,631–0,663)	0,641 (0,626–0,657)	0,616 (0,603–0,628)
Tuntud riskitegurid	0,631 (0,615–0,648)	0,638 (0,622–0,654)	0,597 (0,585–0,610)
Tuntud riskitegurid+PRS	0,695 (0,680–0,710)	0,695 (0,681–0,710)	0,649 (0,636–0,661)

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Hanna Kangur,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Kodade virvendusarütmi ja laperduse riski prognoosimine TÜ Eesti Geenivaramu näitel“, mille juhendajad on Krista Fischer ja Tuuli Puusepp, 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.

Hanna Kangur

12.05.2026