

TARTU ÜLIKOOL

LOODUS- JA TÄPPISTEADUSTE VALDKOND

MATEMAATIKA JA STATISTIKA INSTITUUT

Gailin Savi

**Kardiovaskulaarsete haiguste ja 2. tüüpi  
diabeedi riskimudelite hindamine TÜ Eesti  
geenivaramu andmete põhjal**

Matemaatiline statistika

Bakalaureusetöö (9 EAP)

Juhendajad: PhD Mart Kals

PhD Krista Fischer

TARTU 2025

**Kardiovaskulaarsete haiguste ja 2. tüüpi diabeedi riskimudelite  
hindamine TÜ Eesti geenivaramu andmete põhjal**

Bakalaureusetöö

Gailin Savi

**Lühikokkuvõte** Bakalaureusetöö eesmärk on välja töötada riski ennustavaid mudeleid neljale kardiovaskulaarsele haigusele ja 2. tüüpi diabeedile. Kõigile haigustele on loodud kaks mudelit ja seejärel võrreldud kliiniliselt kasutatava SCORE2 mudeli riskifaktoritega. Esimene loodud mudel sisaldab polügeenseid riskiskoori, metaboolomika näitajaid ja elustiililisi tunnuseid. Teine mudel on sobitatud vaid polügeensete riskiskooride ja vanuse põhjal. Analüüsis on kasutatud Tartu Ülikooli Eesti geenivaramu andmeid. Töö teoreetilises osas antakse ülevaade Eesti geenivaramust, uuritavatest haigustest, metaboliitidest, polügeensetest riskiskooridest, kasutatavatest analüüsimeetoditest ja SCORE2 riskiskoorist. Praktilises osas on kirjeldatud kasutatavaid andmeid ja mudelite tulemusi. Välja töötatud mudelite prognooside täpsusi on võrreldud Harrelli C-indeksi põhjal.

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

**Märksõnad:** kardiovaskulaarne haigus, 2. tüüpi diabeet, metaboliit, polügeenne riskiskoor, elukestusanalüüs, C-indeks.

**Evaluation of risk models for cardiovascular diseases and type 2 diabetes based on data from the Estonian biobank of the University of Tartu**

Bachelor thesis

Gailin Savi

**Abstract** The aim of this bachelor's thesis is to develop risk prediction models for four cardiovascular diseases and type 2 diabetes. For each disease, two models were developed and subsequently compared to the clinically used SCORE2 risk factors. The first model incorporates polygenic risk scores, metabolomic markers and lifestyle factors. The second model is solely based on polygenic risk scores and age. The analysis is based on data from the Estonian biobank of the University of Tartu. The theoretical part of the thesis provides an overview of the Estonian biobank, the diseases under study, metabolites, polygenic risk scores, of used methodology and the SCORE2 risk algorithm. The practical section describes the dataset and model results. The predictive accuracy of the developed models are evaluated using Harrell's C-index.

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

**Key Words:** cardiovascular disease, type 2 diabetes, metabolite, polygenic risk score, survival analysis, C-index.

# Sisukord

<b>Sissejuhatus</b>	<b>6</b>
<b>1 Eesti geenivaramu</b>	<b>7</b>
<b>2 Uuritavad haigused</b>	<b>9</b>
2.1 Kardiovaskulaarne haigus . . . . .	9
2.2 Kodade virvendusarütmia . . . . .	9
2.3 Koronaararterite haigus . . . . .	10
2.4 Südamepuudulikkus . . . . .	11
2.5 Diabeet . . . . .	12
<b>3 Metaboliidid</b>	<b>14</b>
<b>4 Ülegenoomne seoseuuring</b>	<b>16</b>
4.1 Polügeenne riskiskoor . . . . .	16
<b>5 Lassoregressioon</b>	<b>18</b>
<b>6 Coxi võrdeliste riskide mudel</b>	<b>19</b>
6.1 Võrdeliste riskide eelduste kontroll . . . . .	20
<b>7 Harrelli C-indeks</b>	<b>22</b>
<b>8 SCORE2 skoor</b>	<b>23</b>
<b>9 Metoodika</b>	<b>25</b>
9.1 Andmete kirjeldus . . . . .	25

9.1.1	Uuritavate haiguste RHK-10 koodid . . . . .	25
9.2	Riskifaktorid . . . . .	25
9.3	Kaasamiskriteeriumid . . . . .	26
9.4	Mudelid . . . . .	27
<b>10</b>	<b>Analiüs</b>	<b>28</b>
10.1	Kirjeldav statistika . . . . .	28
10.2	Kardiovaskulaarse haiguse riskitegurid . . . . .	33
10.2.1	Vanusel ja PRS-idel põhinevad mudelid . . . . .	35
10.3	Kodade virvendusarütmia riskitegurid . . . . .	36
10.3.1	Vanusel ja PRS-idel põhinevad mudelid . . . . .	38
10.4	Koronaararteri haiguse riskitegurid . . . . .	39
10.4.1	Vanusel ja PRS-idel põhinevad mudelid . . . . .	43
10.5	Südamepuudulikkuse riskitegurid . . . . .	44
10.5.1	Vanusel ja PRS-idel põhinevad mudelid . . . . .	47
10.6	2. tüüpi diabeedi riskitegurid . . . . .	48
10.6.1	Vanusel ja PRS-idel põhinevad mudelid . . . . .	51
10.7	SCORE2 komponentidega mudelid . . . . .	51
10.8	Mudelite võrdlus . . . . .	53
	<b>Kokkuvõte</b>	<b>55</b>
	<b>Kasutatud kirjandus</b>	<b>56</b>

## Kasutatud lühendid

KVH	Kardiovaskulaarne haigus
KVA	Kodade virvendusarütmia
KAH	Koronaararterite haigus
BCAA	Hargnenud ahelaga aminohape
GWAS	Ülegenoomne seoseuring
SNP	Üksiknukleotiidne polümorfism
PRS	Polügeenne riskiskoor
LD	Aheldustasakaalutus
CS	Pidev kahanemine
LASSO	Lassoregressioon
HDL	Kõrge tihedusega lipoproteiin
LDL	Madala tihedusega lipoproteiin
IDL	Keskmise tihedusega lipoproteiin
VLDL	Väga madala tihedusega lipoproteiin

## Sissejuhatus

Eestis on keskmine oodatav eluiga tõusnud, aga tervena elatud aastate arv väheneb ja krooniliste haiguste, sealhulgas kardiovaskulaarsete haiguste ja diabeedi levimus on kõrge [1]. Euroopa Komisjoni ning Majandusliku Koostöö ja Arengu Organisatsiooni 2016. aastal välja antud raportis on esile toodud, et varajane haigusriskide hindamine võimaldab tõhusamat ennetust ja personaalsemat ravi, mis omakorda aitab hoida tervishoiusüsteeme jätkusuutlikena [2]. Käesoleva bakalaureusetöö eesmärk on välja töötada erinevate kardiovaskulaarsete haiguste ja 2. tüüpi diabeedi riskimudelid, kasutades kliinilisi riskifaktoreid, polügeenset riski ja metaboliite. Saadud mudelite prognoosivõimet on võrreldud mudelitega, mis sisaldavad klassikalisi kliinilises praktikas kasutatavaid riskifaktoreid. Töös kasutatud andmed on pärit Tartu Ülikooli Eesti geenivaramust.

Töö jaguneb teoreetiliseks ja praktiliseks osaks. Teoreetiline osa koosneb üheksast peatükist. Esimeses peatükis on tutvustatud Eesti geenivaramut. Teises peatükis kirjeldatakse uuritavaid haigusi ja tuuakse välja nende riskitegurid. Kolmas peatükk annab ülevaate metaboliitidest. Neljandas peatükis on kirjeldatud ülegeenoomset seoseuuringut ja polügeenseid riskiskoore. Viiendas peatükis on ülevaade lassoregressioonmudelist. Kuuendas peatükis kirjeldatakse töös kasutatatud elukesetusanalüüsi meetodit. Seitsmes peatükk sisaldab teooriat mudeli prognoositäpsust kirjeldava C-indeksi kohta. Kaheksandas peatükis on kirjeldatud kardiovaskulaarset riskiskoori SCORE2 ja viimaks on kirjeldatud töö metoodikat.

Praktilises osas kirjeldatakse töös kasutatud tunnuseid koos kirjeldava statistikaga. Ühtlasi on kirjeldatud töö tulemusi. Iga haiguse jaoks on hinnatud kolm mudelit ning võrreldud nende mudelite prognoosivõimeid.

Analüüside läbiviimisel on kasutatud rakendustarkvara R.

# 1 Eesti geenivaramu

Tartu Ülikooli Eesti geenivaramu on riiklik biopank, mis asutati 1999. aastal. Sellest ajast alates on geenivaramu läbinud mitmeid arengufaase. Esimese kümnendi jooksul oli suurem fookus osalejate kaasamisel ja andmete kogumisel. Aastaks 2015 koguti kohort, kuhu kuulus 52 000 indiviidi. Sel ajal osaleti näiteks aktiivselt rahvusvahelistes kogu genoomi hõlmavates assotsiatsioonuuuringutes ehk ülegenoomsetes seoseuuringutes. Kuigi Eesti geenivaramu on märkimisväärselt panustanud ülegenoomsete seoseuuringutega inimgeneetika alastesse avastustesse, siis tagasikutsumisega ehk *recall by genotype* uuringutest saadud tulemused on olnud ühed silmapaistvamatest. Sellist tüüpi uuring on küllaltki unikaalne iseärasus maailma juhtivate biopankade hulgas. Tänu võimalusele võtta ühendust geenidoonoritega, kellel on kindel geneetiline profiil, on koostöös arstidega täpsemalt uuritud mitmete haiguste geneetilist riski, mis võimaldaks paremini tuvastada kõrge riskiga isikuid. [3]

Biopanga tähelepanuväärne töö andis Eesti riigile aluse toetada täiendava 100 000 isiku kaasamist ja genotüpiseerimist aastal 2018 ning lisaks 50 000 inimest 2019. aastal. Tänu digiteenuste arengule sai selles etapis lihtsustada ka geenidoonorite kaasamist. Ühe näitena saab tuua küsimutiku lühendamise: eemaldati varasemate diagnooside ja ravimite kohta käivad blokid. Selline teostus oli võimalik tänu Eesti Tervisekassast, Tervise Infosüsteemi andmebaasidest ja paljudest registritest saadavale informatsioonile. [3]

Tänaseks on biopangaga liitunud umbes 212 000 eestimaalast, mis moodustab ligikaudu viiendiku riigi täisealisest populatsioonist. Genotüpiseeritud on kõigi doonorite DNA-d, on määratud metaboliidid ning andmete vahel on loodud seos erinevate terviseregistritega nagu nt vähiregistriga ja müokardiinfarktiregistriga. [3] Ühtlasi valmis 2024. aastal portaal [www.portaal.geenidoonor.ee](http://www.portaal.geenidoonor.ee), kust geenidoonorid saavad personaalset teavet erinevate haiguste eelsoodumuste ja päritolu kohta. Üks portaali loomise eesmärke oli suunata inimeste tervisekäitumist selliselt,

et vähendada või ennetada nende riski teatud haiguste tekkimisel. [4] Lisaks on ainuüksi Tartu Ülikooli Eesti geenivaramu teaduskeskuse teadlased geenivaramu andmete alusel panustanud üle 300 uuringu valmimisse ning osalenud enam kui 800 publikatsiooni avaldamises [3].

## 2 Uuritavad haigused

### 2.1 Kardiovaskulaarne haigus

Südame-veresoonkonna haiguste ehk kardiovaskulaarsete (KV) haiguste alla kuuluvad haigused, mis puudutavad veresooni või südametööd. Näidetena võib tuua südame isheemiatõve, südamepuudulikkuse, südame rütmihäired ja kõrgvererõhktõve. [5]

KV haigused on peamine surmapõhjus kogu maailmas, põhjustades igal aastal hinnanguliselt pea 18 miljoni inimese surma [6]. Eestis suri 2024. aastal 15 542 inimest ja neist 7242 KV haiguse tõttu, mis näitab, et pea pooled surmad on põhjustatud antud valdkonna haigustest [7].

KV haigustel on mitmeid riskifaktoreid. Näiteks kõrge vererõhk, kõrge kolesterooli tase, 2. tüüpi diabeet, mõni autoimmuunhaigus, krooniline neeruhaigus või KV haiguse perekondlik ajalugu. Lisaks võivad rasedusaegsed haigused nagu preeklampsia ja rasedusaegne diabeet soodustada KV haiguste teket. Ühtlasi mõjutab haiguse teket vähene füüsiline aktiivsus, ülekaalulisus, liigne alkoholarbimine, suitsetamine, retsepti- ja/või meelelahutuslike ravimite väärkasutamine ja kõrge naatriumi-, suhkru- ja rasvasisaldusega dieet. [5] Naatriumi on peaaegu kõikides toitudes, kuid kõige rohkem leidub seda soolas.

### 2.2 Kodade virvendusarütmia

Süda koosneb neljast kambrist: kahest kojast ja kahest vatsakesest. Tavaliselt töötab süda kindla rütmi järgi, mida juhivad siinussõlm, saates elektriimpulsse, mis panevad südame kokkutõmbuma ja verd ringi pumpama. Kodade virvendusarütmia (KVA) korral ei tööta siinussõlm korrapäraselt ning elektriimpulsid liiguvad kodades segaselt, põhjustades südame rütmihäire. Selle tagajärjel on pulss ebaregulaarne ja veri ei liigu südames tõhusalt. KVA võib esineda ajutiste hoogudena või

olla püsiv. Tavaliselt on tegemist kroonilise haigusega, mille episoodid muutuvad aja jooksul sagedasemaks ja püsivamaks. [8]

KVA on kõige sagedasem südame rütmihäire [8], mis esineb viiel inimesel 1000-st [9]. Eestis on haigus diagnoositud umbes 26 000 inimesel [8]. KVA-ga elab 2019. aasta seisuga ligikaudu 59 miljonit inimest, võrreldes 2010. aastaga, kui haigus oli diagnoositud 33,5 miljonil inimesel [10].

KVA esineb sagedamini vanemas eas ning alates 45. eluaastast on haiguse tekkimise risk umbes ühel inimesel 3-5-st [10]. Samas on oluline roll haiguse tekkele ja selle püsijäämisele ka elustiilil. Suitsetamine, liigne alkoholarbimine, vähene füüsiline aktiivsus, tasakaalustamata toitumine ja liigne stress on kõik KVA teket soodustavad riskitegurid. [8]

KVA kõige sagedasemad tüsistused on insult ja südamepuudulikkus. Insuldi tekket risk on KVA diagnoosiga inimesetel umbes viis korda suurem, kui neil, kellel seda haigust pole. Kuigi KVA-d pole võimalik ravida, siis tervisliku elustiili ja arsti poolt määratud ravirežiimi järgimisega on võimalik elada täisväärtuslikku elu. [8]

## 2.3 Koronaarterite haigus

Pärgarterid on veresooned, mis varustavad südant verrega. Pärgarterite lubjastumiseks ehk ateroskleroosiks nimetatakse protsessi, mille käigus vererasvad, kolesterool ja muud ained ladestuvad arterite seintesse, mis lõpuks viib pärgarterite ahenemiseni. Südant toitvate pärgarterite lubjastumise tõttu areneb koronaarterite haigus (KAH), mis on tuntud ka kui isheemiline südamehaigus. [11]

KAH on üks sagedasemalt diagnoositavaid haigusi. Terves maailmas põeb haigust hinnanguliselt rohkem kui 250 miljonil inimest ning ligikaudu üks seitsmest surmast maailmas on põhjustatud KAH tõttu. [12]

Vananemisega suureneb arterite ahenemise risk, mistõttu suureneb risk KAH-sse haigestuda. Ühtlasi on meestel suurem risk, kuid naistel suureneb haiguse tekkerisk

pärast menopausi. Samuti on KAH tugeva geneetilise eelsoodumusega. Samas on ka hulgaliselt riskifaktoreid, mis on seotud elustiiliga: suitsetamine, kolesterooli tase, diabeet, ülekaalulisus, krooniline neeruhaigus, vähene füüsiline aktiivsus, liigne stress, tasakaalustamata toitumine, liigne alkoholi tarbimine ja liigne või vähene magamine. [11]

KAH võimalikud tüsistused on stenokardia ehk rinnaangiin, südameinfarkt, südamepuudulikkus ja arütmia. Ravimid ja ka mõningad operatsioonid võivad haigust leevendada, kuid elustiili muutused võivad ära hoida haiguse tekke ja aidata ka võimalike tüsistuste eest. [11]

## 2.4 Südamepuudulikkus

Südamepuudulikkuse puhul ei suuda süda verd pumbata vajalikus mahus ning selle tulemusena ei jõua hapnikurikas veri organismi. Südamepuudulikkus ei ole iseseisvalt tekkiv haigus, s.t see tekib teiste diagnoosimata südamehaiguste tagajärjel. [13] Olenevalt haiguse avaldumise kiirusele eristatakse ägedat ja kroonilist südamepuudulikkust. Ägeda südamepuudulikkuse puhul on see avaldanud mõne teise ägeda südamehaiguse tagajärjel, nt südamelihase infarkt. [14] Öeldakse, et südamepuudulikkus on krooniline, kui see areneb aeglaselt elu jooksul [13].

Järgnevalt on nimetatud peamised südamepuudulikkuse põhjused: südame isheemiatõbi, südamelihase infarkt, kõrgvererõhktõbi, müokardiit, arütmia, diabeet, südameklapi rike ja kaasasündinud südamerike. Lisaks nendele võib äkilise südamepuudulikkuse tekitada ka allergilised reaktsioonid ja trombid kopsudes. Ühtlasi on selle haiguse riskifaktorid ka uneapnea, ülekaalulisus, vanus, liigne alkoholi tarvitamine ja suitsetamine. Südamepuudulikkuse võimalikud tüsistused on neeruhaigus või puudulikkus, maksakahjustus, muud muutused südames ja südame äkksurm. [15]

Nii USA-s kui ka Euroopas põeb südamepuudulikkust 1-2% elanikest, kuid 70-aastaste ja vanemate seas esineb südamepuudulikkus rohkem kui 10%-l. Kuigi Eesti

populatsioonis südamepuudulikkuse kohta käiv täpne statistika puudub, siis eelnevast lähtudes võib südamepuudulikkusega isikute arv olla ligikaudu 25 000. [16] Euroopas elab südamepuudulikkusega ligikaudu 15 miljonit inimest [17].

## 2.5 Diabeet

Diabeet ehk suhkruhaigus on energiaainevahetuse püsiv häire, mille korral veresuhkrutase on liiga kõrge ning kõhunäär ei tooda piisavalt insuliini. Insuliin on hormoon, mis aitab glükoosil jõuda rakkudeni ning mille läbi organism saab energiat toitainete omastamise kaudu. Eristatakse kolme tüüpi diabeeti: 1. ja 2. tüüpi ning rasedusaegset diabeeti ehk gestatsiooni. 1. tüüpi diabeedi korral ei tooda kõhunääre insuliini, sest immuunsüsteem ründab insuliini tootvaid rakke. Seda tüüpi diabeeti haigestutakse tavaliselt lapseas või noores täiskasvanueas. 2. tüüpi diabeedi puhul ei tooda kõhunäär piisavalt insuliini ja seega ei suuda organism veresuhkrutaset ise reguleerida. Seda tüüpi diabeedi tekkerisk on suurem inimestel, kellel pereliige põeb haigust või kes on ülekaaluline. Gestatsioon on raseduaegne diabeet, mis tavaliselt kaob pärast lapse sündi. Küll aga on gestatsiooni põdevatel naistel suurenenud risk 1. ja 2. tüüpi diabeedi arenemise suhtes. [18]

Eestis on diabeet diagnoositud ligikaudu 70 000 inimesel. Neist 7000-l on 1. tüüpi diabeet ja umbes 65 000-l 2. tüüpi diabeet. [19] Maailmas põeb diabeeti umbes 830 miljonit inimest ning diabeedikute arv on olnud pidevas kasvamises viimase kümnendi jooksul [20].

1. tüüpi diabeeti tekitavad riskifaktorid on vanus, geneetiline eelsoodumus ja keskkonnategurid. 2. tüüpi diabeedi riskitegureid on märksa rohkem, kuid paljusi neist on igal inimesel võimalik kontrollida ja seega ka haiguse tekkimist kas ennetada või edasi lükata ning haiguse tekkimisel seda kontrolli all hoida. Haigusrisi mõjutab pärilikkus, ülekaalulisus, rasvtõbi, metaboolne sündroom, podagra, mõningate ravimite tarvitamine (kortisoon, beetablokkerid), vähene füüsiline aktiivsus, liigne alkoholi tarbimine, suitsetamine, halb une kvaliteet, stress ja rasedusaegne diabeet.

[18]

Diabeedi puhul on seega oluline hoida veresuhkrutase normi piires, vastasel juhul on organism avatud mitmete tüsistuste tekkele. Diabeediga seonduvad tüsistused on diabeetiline ketoatsidoos, retinopaatia, nefropaatia, neuropaatia, periodontiit, diabeetiline jalg ja mitmesugused südame-veresoonkonna haigused. [18]

### 3 Metaboliidid

Metaboliidid on organismi ainevahetuse protsessi tulemusena moodustunud ained [21]. Metaboliidid on bioloogiliste protsesside ja haigusseisundite põhinäitajad. Ühtlasi võimaldavad need paremini mõista inimese molekulaarbioloogiat ning selalased teadustööd võivad avaldada seoseid geenide ja metaboolsete radade vahel. [22] Metaboliitide andmete kättesaamiseks saab kasutada tuumamagnetresonantsi spektroskoopiat ning Eesti biopangas kasutatakse selleks Soome ettevõtte Nightingale Health platvormi, mis mõõdab lipiide, aminohappeid, rasvhappeid jne [23]. Ühendkuningriigi biopanga ja Eesti geenivaramu andmete põhjal uuriti 249 metaboliidi mõju südame isheemiatõvele ja 2. tüüpi diabeedile. Analüüsis kasutati genoomiülest Mendeli randomiseerimist, mis näitas, et 85% metaboliitidest on seotud südame isheemiatõve riskiga ning 63% seotud 2. tüüpi diabeediga, kuid suur osa nendest seostest võivad olla pleiotroopsed ehk sama metaboliit mõjutab korraga mitme tunnuse kujunemist. Ühtlasi leiti, et hargnenud ahelaga aminohappe (*Branched chain amino acids*, BCAA) taseme alandamine ravimitega ei pruugi vähendada 2. tüüpi diabeedi riski, mida varem oli arvatud. [22]

Metaboolsed riskiskoorid põhinevad vereproovist määratud biomarkerite efektide summeerimisel. On näidatud, et metaboolseid riskiskoore kasutades on võimalik prognoosida mitmeid levinumaid haiguseid, sealhulgas südame-veresoonkonna haiguseid ja 2. tüüpi diabeeti. Metaboolne riskiskoor erineb polügeensest riskiskoorist, sest see võib aja jooksul elustiili ja ravi tõttu muutuda, mis võimaldab jälgida muutusi isikute riskiprofiilides. [23]

Ühendkuningriigi, Eesti ja Soome biopankade andmete põhjal tehtud uuringus analüüsiti omavahel metaboolseid, polügeenseid ja kliinilisi riskiskoore 12-ne haiguse puhul. Analüüsi tulemusena leiti, et metaboolsed riskiskoorid on enamike uuritavate haiguste puhul haiguse tekkega tugevamalt seotud kui polügeensed riskiskoorid. Samas enamasti andis parima tulemuse mudel, mis sisaldas nii polügeenset kui ka metaboolset riskiskoori, mis näitab, et need sisaldavad vähemalt osaliselt erinevat

teavet. Ühtlasi leiti, et metaboolsed riskiskoorid parandavad prognoosivõimet, kui neid kombineerida kliiniliste riskiteguritega. [23]

## 4 Ülegenoomne seoseuuring

Ülegenoomsete seoseuuringute (*Genome-wide association study*, GWAS) mõiste esines esmakordselt 2007. aastal. Tänapäevaks on sellest saanud üks populaarsemaid uuringuliike, mida geneetilise epidemoloogia valdkonnas kasutatakse. [24] Ülegenoomse seoseuuringu peamine eesmärk on leida geenivariante, mis seostuvad erinevate haiguste või tunnustega. Saadud tulemused võimaldavad näiteks tuvastada, kui suur on haiguse tekkerisk erinevate geenivariantidega inimestel. Ühtlasi on tulemuste abil võimalik välja töötada ennetusviise ja ravivõimalusi. [25]

### 4.1 Polügeenne riskiskoor

Üksiknukleotiidne polümorfism (*Single nucleotide polymorphism*, SNP) on DNA järjestuse variatsioon [25], mis on toimunud ühe genoomi nukleotiidi muutumisel. Polügeenne riskiskoor (*Polygenic risk score*, PRS) on genoomis esinevate üksikute SNP-de põhjal hinnatud arvutuslik näitaja, mis iseloomustab geneetilist eelsoodumust uuritavatele tunnustele. Esimeste PRS-ide arvutamisel kasutati vaid selliseid SNP-sid, mis olid statistiliselt olulised ülegenoomsetes seoseuuringutes. See lähenemine ei osutunud aga kuigi edukaks, kuna saadud prognoosivõime jäi alla lähisugulastelt saadud haigusloole. Põhjus seisnes selles, et pärast mitmese testimise korrektsiooni jäid vähesed SNP-d statistiliselt oluliseks. Tänapäevaks on aga leitud, et prognoosivõime suureneb, kui kaasata suurem hulk SNP-sid, k.a sellised, mille puudub oluline mõju uuritavale tunnusele. [24] Selleks on välja töötatud mitmeid meetodeid.

Üks lihtne ja laialt levinud meetod on P+T (*Pruning and Thresholding*), mille käigus tehakse esmalt aheldustasakaalutuse (*Linkage disequilibrium*, LD) põhine valik sõltumatute SNP-de kaasamiseks [26], s.t eemaldatakse lähestikku paiknevad ja tugevas korrelatsioonis olevad geenivariandid [25]. Seejärel rakendatakse p-väärtuse lävendit, et kaasata SNP-d vastavalt GWAS tulemustele [26].

Teine PRS arvutusmeetod, PRS-CS, põhineb Bayesi regressioonil, mis kasutab SNP-de kaalude leidmiseks pideva kahanemise (*continuous shrinkage*, CS) eeljaotust. CS-i abil saab otsustada, millise SNP mõju vähendada, sõltuvalt sellele, kui tugevalt on marker seotud uuritava tunnusega. Lisaks kasutatakse LD-d, mille abil kohandatakse markeri mõju tunnusele. [27]

Olgu  $k$  SNP-de arv,  $\beta_j$   $j$ -nda SNP mõju suurus ehk kaal ning  $X_j$   $j$ -nda SNP efekialleelide arv. Efekialleel on SNP alleel, mille mõju tunnusele soovitakse uurida. SNP-d on dialleelsed, seega on  $X_j$  võimalikud väärtused hulgast  $\{0, 1, 2\}$ . [24] Tegelikuses võivad  $X_j$  väärtused olla vahemikust  $[0, 2]$ , kuna sageli kasutatakse imputeerimist. Imputeerimise vajadus tuleneb asjaolust, et SNP-de arv on tohutu suur ja kõigi genotüpiseerimine ei ole kulutõhus. Selleks valitakse välja SNP-de alamhulk, mis esindab suuremat osa SNP-dest. Genotüpiseerimata SNP-d imputeeritakse mõõdetud SNP-de ja LD alusel. [28]

Lineaarset regressioonimudelit kasutatakse, kui uuritav tunnus ülegenoomses uurin-gus on pidev, kuid haiguse esinemise tõenäosust hinnatakse logistilise regressioon-mudeli abil. Logistilise regressioonimudeli kuju:

$$\ln\left(\frac{p}{1-p}\right) = \mu + l(Z) + \beta_j X_j, \quad (1)$$

kus  $\mu$  on konstant,  $l(Z)$  teised uuritava tunnuse esinemist mõjutavad tegureid,  $X_j$   $j$ -nda SNP efekialleelide arv ja  $\beta_j$  selle kordaja,  $j = 1, \dots, k$ . [24]

Polügeenne riskiskoor  $i$ -ndale indiviidile avaldub kujul:

$$PRS_i = \sum_{j=1}^k w_j X_{ij}, \quad (2)$$

kus  $w_j$  on hinnang  $i$ -nda isiku  $j$ -nda SNP kaalule [28].

## 5 Lassoregressioon

Lassoregressioon (*Least absolute shrinkage and selection operator*, LASSO) on laialdaselt kasutatav statistilise modelleerimise meetod. Selle peamine eesmärk on leida tasakaal mudeli lihtsuse ja täpsuse vahel, seda eriti olukordades, kus argumenttunnuseid on palju. Ühtlasi on meetodi kasutamine kasulik olukordades, kus tunnuste vahel võib esineda tugevat multikollineaarsust. Lassoregressiooni abil on võimalik läbi viia automaatne tunnuste valik ja selleks saab kasutada regulariseerimisparameetrit  $\lambda$ , millega osad mudeli kordajad määratakse nulliks. [29]

Lassoregressioon minimiseerib järgnevat avaldist:

$$RSS + \lambda \sum_{j=1}^p |\beta_j|, \quad (3)$$

kus RSS on jääkruutude summa,  $\beta_j$  on tunnuste kordajad. Regulariseerimisparameeter  $\lambda$  määrab, kui suures ulatuses mudelit lihtsustatakse. Suurem  $\lambda$  tähendab, et rohkemate tunnuste kordajaid määratakse nulliks ja jäetakse seega tunnuste kogust välja. [29]

Parima  $\lambda$  väärtuse valimisel kasutatakse tavaliselt ristvalideerimist [29]. Seda meetodit on kasutatud ka käesolevas bakalaureusetöös.

## 6 Coxi võrdeliste riskide mudel

Järgnev peatükk ja alapeatükk tugineb David Colleti raamatule "Modelling survival data in medical research"[30].

Coxi võrdeliste riskide mudelit saab kasutada, et uurida elukestuse ja selgitavate tunnuste vahelisi seoseid. Elukestusanalüüsis on tähtsal kohal indiviidi suremuse risk igal ajal pärast uuringu algusaega.

Olgu andmed  $n$  indiviidi jaoks ja  $p$  argumenttunnust  $X_1, X_2, \dots, X_p$ , on saadud uuringu algusajal ning need ei muutu ajas. Vektoriga  $\mathbf{x} = (x_1, x_2, \dots, x_p)'$  tähistame argumenttunnuste väärtuseid ning vektor  $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})'$  sisaldab  $i$ -nda indiviidi argumenttunnuste väärtuseid. Olgu  $h_0(t)$  indiviidi riskifunktsioon, kelle kõikide argumenttunnuste väärtused vektoris  $\mathbf{x}$  on nullid. Sellist riskifunktsiooni nimetatakse baasriskifunktsiooniks. Riskifunktsiooni  $i$ -ndale indiviidile saab esitada kujul

$$h_i(t) = \psi(\mathbf{x}_i)h_0(t), \quad (4)$$

kus  $\psi(\mathbf{x}_i)$  on muutuja  $\mathbf{x}_i$  funktsioon.

Funktsioon  $\psi(\mathbf{x}_i)$  näitab ajahetkel  $t$   $i$ -nda indiviidi suhtelist riski võrreldes indiviidiga, kelle kõik argumenttunnuste väärtused on nullid. Suhteline risk  $\psi(\mathbf{x}_i)$  ei saa olla negatiivne, seega sobib kasutada eksponentfunktsiooni kujul  $e^{\eta_i}$ , kus  $\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} = \boldsymbol{\beta}' \mathbf{x}_i$  ja vektor  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$  tähistab muutujate kordajaid. Sellisel juhul saab riskifunktsiooni avaldada kujul:

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

Eelnevat riskifunktsiooni, mille puhul ei tehta eeldusi baasriskifunktsiooni kohta ning mis teeb mudeli poolparameetriliseks, nimetatakse Coxi võrdeliste riskide mudeliks. Antud mudeli puhul hinnatakse muutujate kordajaid suurima tõepära meetodi abil.

Olgu  $r$  indiviidil  $n$ -st sündmus toimunud, kusjuures neil kõigil on sündmuse toimumishetk erinev. Järelikult on  $n-r$  indiviidi paremalt tsenseeritud, mis tähendab, et nende puhul toimub sündmus (kui üldse) pärast jälgimisperioodi lõppu. Toimumishetked järjestame vastavalt  $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ . Tähistuse  $R(t_{(j)})$  all mõistame ajahetkel  $t_{(j)}$  riskigrupi kuuluvaid isikuid. On näidatud, et mudelile (5) sobiva tõepärafunktsiooni kuju on

$$L(\boldsymbol{\beta}) = \prod_{j=1}^r \frac{e^{\boldsymbol{\beta}' \mathbf{x}_{(j)}}}{\sum_{l \in R(t_{(j)})} e^{\boldsymbol{\beta}' \mathbf{x}_l}}, \quad (6)$$

kus vektor  $\mathbf{x}_{(j)}$  tähistab indiviidi tunnuse väärtuseid, kellel toimus sündmus  $j$ -ndal ajahetkel ning murrualune summa summeerib  $e^{\boldsymbol{\beta}' \mathbf{x}_l}$  nende väärtused, kes kuuluvad riskigrupi ajahetkel  $t_{(j)}$ . Antud tõepärafunktsioon sõltub sündmuse toimumishetkede järjestusest, mitte täpsetest aegadest ja seepärast on tegemist osalise tõepärafunktsiooniga. Argumenttunnuste kordajate väärtuseid saab hinnata tõepärafunktsiooni (6) maksimeerimise abil kasutades numbrilisi meetodeid.

## 6.1 Võrdeliste riskide eelduste kontroll

Võrdeliste riskide kontrollimiseks on mitmeid meetodeid. Järgnevalt on tutvustatud Scoenfeldi jääkide analüüsi meetodit, mis erineb teistest meetoditest selle poolest, et iga indiviidi jaoks on iga mudelisse kaasatava tunnuse kohta jääkide väärtuste kogum.

Olgu vaatluse all  $n$  indiviidi ja neist  $r$  isikul sündmus toimunud, seega  $n-r$  indiviidi on paremalt tsenseeritud. Igal indiviidil olgu mõõdetud  $p$  tunnust  $X_1, X_2, \dots, X_p$  ning Coxi võrdeliste riskide mudeli riskifunktsiooni parameetrite  $\boldsymbol{\beta}$  hinnangud  $\hat{\boldsymbol{\beta}}$  leitud. Seega  $i$ -nda indiviidi riskifunktsiooni hinnang on kujul:

$$\hat{h}_i(t) = e^{\hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_p x_{pi}} \hat{h}_0(t) = e^{\hat{\boldsymbol{\beta}}' \mathbf{x}_i} \hat{h}_0(t), \quad (7)$$

kus  $\hat{\boldsymbol{\beta}}' \mathbf{x}_i$  on  $i$ -nda indiviidi riskiskoor ja  $\hat{h}_0(t)$  on baasriskifunktsiooni hinnang.

Sümboliga  $\delta$  tähistatakse indikaatortunnust, mille väärtus on 1, kui indiviidil sündmus toimus, vastasel juhul on väärtus 0. Scoenfeldi jääk  $i$ -nda indiviidi ja  $j$ -nda tunnuse  $X_j$  jaoks on leitav valemist:

$$r_{Sji} = \delta_i \{x_{ji} - \hat{a}_{ji}\}, \quad (8)$$

kus  $x_{ji}$  on  $j$ -nda tunnuse väärtus  $i$ -ndal indiviidil. Eelnevas valemis olev suurus  $\hat{a}_{ji}$  esitub kujul:

$$\hat{a}_{ji} = \frac{\sum_{l \in R(t_i)} x_{jl} e^{\hat{\beta}' \mathbf{x}_l}}{\sum_{l \in R(t_i)} e^{\hat{\beta}' \mathbf{x}_l}}, \quad (9)$$

kus  $R(t_i)$  tähistab ajahetkel  $t_i$  riskigruppi kuuluvaid isikuid.

Olgu  $i$ -nda indiviidi Scoenfeldi jääkide vektor  $\mathbf{r}_{Si} = (r_{S1i}, r_{S2i}, \dots, r_{Spi})'$  ning  $\text{var}(\hat{\beta})$  hinnatud parameetrite  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$  kovaratsioonmaatriks. Kaalutud Scoenfeldi jäägid on leitavad valemiga

$$\mathbf{r}_{Si}^* = d \text{var}(\hat{\beta}) \mathbf{r}_{Si}, \quad (10)$$

kus  $d$  on arv, mis näitab, kui mitmel isikul sündmus toimus.

Scoenfeldi jääkide puhul kehtib omadus  $E(r_{Sji}^*) \approx \beta_j(t_i) - \hat{\beta}_j$ , kus  $\beta_j(t)$  on tunnuse  $X_j$  ajast sõltuv kordaja ja  $\beta_j(t_i)$  on kordaja väärtus ajahetkel  $t_i$ . Kujutades graafikul väärtuseid  $r_{Sji}^* + \hat{\beta}_j$ , saab kontrollida võrdelise suhte eelduse kehtivust tunnuse  $X_j$  puhul. Nende väärtuste paiknemine horisontaaljoonel viitab eelduse kehtimisele.

## 7 Harrelli C-indeks

Harrelli C-indeks (*concordance index*) on elukestusanalüüsis laialdaselt kasutatav mõõdik, millega hinnatakse, kui hästi suudab mudel järjestada indiviidid nende sündmuse aja põhjal. C-indeksi eelis võrreldes teiste mudeli headuse mõõdikutega, on see, et indeks arvestab andmete võimalikku tsenseeritust (s.t sündmuse mitte-toimumist jälgimisperioodil). [31]

C-indeks näitab isikutele hinnatud riskiskooride ja tegeliku sündmuse järjekorra vastavust. Olgu vaatluse all kaks indiviidi  $i$  ja  $j$  ning neil mõlemal on jälgimisperioodi jooksul sündmus toimunud. Öeldakse, et tegu on kooskõlalise indiviidide paariga, kui  $i$ -nda indiviidi riskiskoor on kõrgem ja tal toimus sündmus enne kui  $j$ -ndal isikul. Selliste eelduste põhjal öeldakse, et paar on ebakõlaline, kui kehtib vastupidine (nt  $i$ -nda indiviidi riskiskoor on kõrgem, aga sündmus toimus enne  $j$ -ndal indiviidil). Kooskõlaliseks loetakse ka seda paari, mille korral  $i$ -ndal indiviidil on sündmus toimunud ja tema riskiskoor on kõrgem kui  $j$ -ndal isikul, kuid viimasel pole sündmust jälgimisperioodil toimunud. Taaskord loetakse sellisel juhul vastupidist ebakõlaliseks paariks. Neid indiviidide paare, kus kummalgi sündmus ei toimunud, ei arvestata C-indeksi arvutamisel. C-indeksit arvutatakse järgmise valemi abil [31]:

$$c = \frac{\text{kooskõlaliste paaride arv}}{\text{kooskõlaliste paaride arv} + \text{ebakooskõlaliste paaride arv}}. \quad (11)$$

C-indeksi väärtused on vahemikus  $[0, 1]$ . Väärtuse 1 korral prognoosib mudel täpselt, kellel indiviididest sündmus esimesena toimub. C-indeks väärtusega 0 näitab, et mudel järjestab kõik indiviidide paarid vastupidiselt, s.t igas paaris omistatakse kõrgem riskiskoor indiviidile, kellel toimus sündmus hiljem. Mudel, mille C-indeks on 0,5, ei erista sündmuse varasema ja hilisema toimumisega isikuid paremini, kui juhuslik järjestamine. [31]

## 8 SCORE2 skoor

Välja on töötatud mitmeid erinevaid kardiovaskulaarseid riskiskooore. Näiteks 2021. aastal loodud SCORE2 (*Systemic coronary risk evaluation 2*), mis töötati välja eesmärgiga pakkuda täpsemat ja ajakohastatud hinnangut 10-aastasele KV haiguse tekkeriskile inimestel vanuses 40-69 aastat. Mudeli abil saab prognoosida esmast haigestumist inimestel, kellel ei ole ühtegi varasemat KV haigust ega diabeeti. Arendatud on ka mudel SCORE2-OP, mis on mõeldud hindamaks riski 70-aastatel ja vanematel indiviididel. SCORE2 on edasiarendus mudelist SCORE ning täiustab eelnevat mudelit mitmetes aspektides. [32]

Mudeli arendamisel kasutati ulatuslikku Euroopa kohortuuringute andmestikku, mis võimaldas arvestada erinevate geograafiliste piirkondade ja riskitasemetega. SCORE2 eristub eelkäijast, SCORE mudelist, selle poolest, et lisaks surmavatele sündmustele võtab see arvesse ka mitteletaalseid juhtumeid, pakkudes seeläbi paremat riskiarvestust. Lisaks on mudel kalibreeritud vastavalt neljale Euroopa riskipiirkonnale: madal, mõõdukas, kõrge ja väga kõrge. [32]

Võrreldes SCORE mudeliga suudab SCORE2 paremini hinnata KV haiguse riski, seda eriti noorematel. Ühtlasi on SCORE2 võimeline paremini eristama kõrge riskiga grupis olevaid inimesi. Mudelit kasutatakse laialdaselt kliinilises praktikas, et tuvastada kõrge ja väga kõrge riskiga isikuid ning määrata neile sobivaid ennetustegevusi ja ravi. Riskitegurid, mida SCORE2 mudel prognoosimisel kasutab, on järgmised:

- vanus,
- sugu,
- süstoolne vererõhk,
- üldkolesterool,
- kõrge tihedusega lipoproteiinide (*High density lipoprotein*, HDL) kolesterool,

- suitsetamise staatus. [32]

## 9 Metoodika

### 9.1 Andmete kirjeldus

Antud töös on kasutatud Tartu Ülikooli Eesti geenivaramu andmeid. Töös on kasutatud elektroonilisi terviseandmeid, metaboliite, PRS-e ja liitumisel täidetud küsimustike andmeid.

#### 9.1.1 Uuritavate haiguste RHK-10 koodid

Uuritavate haiguste määramisel on kasutatud järgmisi rahvusvahelise haiguse klassifikatsioone ehk RHK-10 koode:

- KVH G45, I20-I25, I63 ja I64;
- KVA I48;
- KAH I21-I23, I24.1 ja I25.2;
- Südamepuudulikkus I11.0, I13.0, I13.2, I25.5, I42.0, I42.5, I42.8, I42.9, I50.0 ja I50.9;
- 2. tüüpi diabeet E11.

### 9.2 Riskifaktorid

Töös on kasutatud riskifaktoritena järgmiseid tunnuseid: vanus geenivaramuga liitumisel, sugu, liitumise aasta, surma aasta, diagnoos ja diagnoosi kuupäev. Lisaks 249 metaboliiti ja 12 PRS-i. PRS-ide arvutamisel kasutati kaale PGS kataloogist järgmistele tunnustele: KV haigus (PGS002316), 2. tüüpi diabeet (PGS001781), KVA (PGS004510), südamepuudulikkuse (PGS004532), KAH (PGS003727), hüpertensioon (PGS002701), kehamassiindkes (PGS002844), üldkolesterool (PGS003143),

HDL-kolesterool (PGS002958), madala tihedusega lipoproteiinide (*Low density lipoprotein*, LDL) kolesterool (PGS003033), triklütseriidid (PGS003148) ja glükohe-moglobiin (PGS002697).

Lisaks nendele on kasutatud tunnuseid, mis on saadud geenidoonorite poolt täide-tud küsimustikest. Neist kategoorilised tunnused on järgmised: õppimine (0: ei õpi, 1: õpib), kõrgeim omandatud haridus (põhiharidus, keskharidus, kõrgem haridus), kolme kuu jooksul töötamine (0: ei ole töötanud kolme kuu jooksul, 1: on tööta-nud kolme kuu jooksul), suitsetamise staatus (0: ei suitseta, 1: suitsetab), alkoholi tarbimine (0: ei tarbi, 1: tarbib). Arvulised tunnused olid pikkus, kaal, kehamassi-indeks, süstoolne ja diastoolne vererõhk. Naistel on lisatud veel kolm tunnust: on olnud lapseootel (0: ei ole kunagi olnud, 1: on olnud), hormonaalsete rasestumis-vastaste vahendite kasutamine (0: ei ole kasutanud, 1: on kasutanud) ja raseduste arv.

Jälgimisaeg on arvatud kasutades haigestumise kuupäeva, surma aastat ja lii-tumisaastat. See on leitud 2025. aasta ja liitumisaasta või haigestumisaasta vahena. Juhul, kui indiviid oli surnud, siis surma aasta ja liitumisaasta vahena.

### 9.3 Kaasamiskriteeriumid

Analüüsis on kasutatud geenidoonorite andmeid, kes liitusid geenivaramuga enne 2011. aastat ning olid liitumisaajal 30-70-aastased. Sellele kriteeriumile vastas 34 045 inimest. Seejärel on analüüsist välja jäetud need, kellel puudusid metaboliitide ja PRS-ide andmed, mille tulemusena jäi lõplikusse valimisse 25 197 inimest.

Küsimustike tunnustes esines puuduvaid väärtuseid ning selleks, et vähendada and-mekadu, on kasutatud imputeerimist paketi *mice* abil. Imputeerimist vajasisid aga vähesed andmed. Pea kõigi tunnuste puhul oli puuduvaid väärtusi vähem kui 2%. Kõige enam puuduvaid väärtusi oli tunnusel raseduste arvu (naistel 5,4%).

Indiviidil võis esineda mitu diagnoosi, seega kaasati ta iga haiguse analüüsi selle

konkreetsel haiguse esmasel diagnoosimisel. Kui indiviidil oli haigus diagnoositud enne liitumist või 30 päeva jooksul pärast liitumist, on ta analüüsis välja jäetud. Ühtlasi jäeti välja need isikud, kelle jälgimisperiood oli null aastat ehk need geenidoonorid, kes olid liitunud geenivaramuga ja surnud samal aastal. Lisaks, kui doonori surma põhjuseks oli mõni vereringeelundite haiguse (RHK-10 I kood), loeti ta KV haigusega indiviidiks v.a 2. tüüpi diabeedi puhul.

## 9.4 Mudelid

Iga uuritava haiguse jaoks on hinnatud kolm mudelit. Esimeses mudelis on tunnused valitud LASSO-ga ning seejärel neid kasutatud Coxi võrdeliste riskide mudelis. Teises, Coxi mudelis, on kasutatud vanust ja 12 PRS-i. Kolmas on Coxi võrdeliste riskide mudel, mis sisaldab SCORE2 skoori tunnuseid.

Mudelid on hinnatud meestele ja naistele eraldi. Iga uuritava haiguse puhul on tunnused esimesse mudelisse valitud R paketi *glmnet* funktsiooni *cv.glmnet* abil. Funktsioon *cv.glmnet* kasutab lassoregressiooni, mis teeb automaatse tunnuste valiku. Valitud tunnuseid on seejärel kasutatud Coxi võrdeliste riskide mudelis, kasutades paketi *survival* funktsiooni *coxph*. Coxi võrdeliste riskide mudelis on lisaks LASSO poolt välja valitud tunnustele lisatud ka kõrgeima omandatud hariduse tunnus. Ükshaaval on tunnuseid suurima p-väärtuse põhised välja jäetud, kuni kõik tunnused mudelis olid statistiliselt olulised. Olulisuse nivoona on antud töös kasutatud väärtust 0,05.

## 10 Analüüs

### 10.1 Kirjeldav statistika

Kardiovaskulaarse haiguse mudeldamisel oli andmestikus kokku 21 377 geenidonorit. Kodade virvendusarütmia puhul koosnes andmestik 24 266-st, koronaararteri haiguse andmestik 24 4199-st, südamepuudulikkuse andmestik 22 202-st ja 2. tüüpi diabeedi andmestik 23 610-st geenidonorist. Täpsem soopõhine jaotus haigete ja tervete indiviidide vahel on välja toodud tabelis 1.

Tabel 1: Haigete ja tervete geenidonorite jaotus haiguste ja soo lõikes.

Haigus	Mehed		Naised	
	Haiged	Terved	Haiged	Terved
Kardiovaskulaarne haigus	1832	4322	3339	11 853
Kodade virvendusarütmia	1455	5821	2011	14 979
Koronaararteri haigus	1063	6101	987	16 268
Südamepuudulikkus	1810	4766	3253	12 373
2. tüüpi diabeet	990	6124	1981	14 515

Tabeli 2 põhjal oli kardiovaskulaarse haigusega meestel keskmine haigestumisvanus 52 aastat ja naistel 54 aastat. Haigestunud geenidonoritel oli kõrgem süstoolne ja diastoolne vererõhk võrreldes tervetega, kusjuures erinevus ulatus meestel 9 mmHg ja naistel 9 mmHg süstoolse vererõhu puhul. Alkoholi tarbivate indiviidide osakaal oli väga kõrge kõigis gruppides, mis võib tuleneda tunnuse binaarsusest.

Tabel 2: Kardiovaskulaarse haiguse valitud tunnuste osakaalud ja keskmised haigetel ning tervetel geenidonoritel soo lõikes.

	Mehed		Naised	
	Haiged	Terved	Haiged	Terved
Suitsetajad	66,6%	61,9%	33,0%	36,6%
Alkoholi tarvitajad	94,8%	96,1%	80,9%	88,5%
Põhiharidus	3,8%	1,3%	2,3%	0,8%
Keskharidus	40,9%	36,7%	36,3%	28,6%
Kõrgem haridus	55,3%	62,0%	61,4%	70,6%
Keskmine vanus haigestumisel	52	-	54	-
Keskmine süstoolne vererõhk	136	127	132	123
Keskmine diastoolne vererõhk	84	81	81	77
Keskmine jälgimisperiood	15,8	16,4	17,0	17,1

Tabelist 3 ilmneb, et kodade virvendusarütmia haigestumise keskmine vanus oli meestel 56 ja naistel 58 aastat, mis on pisut kõrgem, kui teiste vaadeldud haiguste puhul. Haigestunutel oli kõrgem vererõhk ja kõrgema haridusega geenidonoreid oli samuti märkimisväärselt rohkem tervete hulgas, mis võib viidata haridustaseme kaitsvatele teguritele tervise suhtes.

Tabel 3: Kodade virvendusarütmia valitud tunnuste osakaalud ja keskmised haigetel ning tervetel geenidonoritel soo lõikes.

	Mehed		Naised	
	Haiged	Terved	Haiged	Terved
Suitsetajad	68,3%	63,5%	29,2%	35,7%
Alkoholi tarvitajad	95,1%	96,0%	78,6%	87,3%
Põhiharidus	4,4%	1,7%	4,3%	0,9%
Keskharidus	42,9%	37,3%	41,8%	30,0%
Kõrgem haridus	52,7%	61,0%	53,9%	69,1%
Keskmine vanus haigestumisel	56	-	58	-
Keskmine süstoolne vererõhk	138	131	136	124
Keskmine diastoolne vererõhk	84	81	82	78
Keskmine jälgimisperiod	14,0	16,3	15,9	17,1

Tabeli 4 põhjal haigestusid mehed koronaararteri haigusesse keskmiselt 56-aastaselt ja naised 60-aastaselt. Sarnaselt teistele südamehaigustele oli haigetel kõrgem vererõhk. Meeste seas oli märgatav suurem suitsetajate osakaal haigestunute hulgas, mis sobitub varasema teaduskirjandusega suitsetamise kui riskifaktori rollist. Haridustasemete võrdluses oli kõrgema haridusega geenidonorite osakaal tunduvalt suurem tervete seas.

Tabel 4: Koronaararteri haiguse valitud tunnuste osakaalud ja keskmised haigetel ja tervetel geenidonoritel soo lõikes.

	Mehed		Naised	
	Haiged	Terved	Haiged	Terved
Suitsetajad	68,1%	62,7%	33,2%	34,6%
Alkoholi tarvitajad	95%	96,0%	76,3%	86,9%
Põhiharidus	5,9%	1,8%	6,3%	1,0%
Keskharidus	44,6%	37,3%	46,2%	30,7%
Kõrgem haridus	49,5%	60,9%	47,5%	68,3%
Keskmine vanus haigestumisel	56	-	60	-
Keskmine süstoolne vererõhk	139	131	139	125
Keskmine diastoolne vererõhk	84	82	83	78
Keskmine jälgimisperiood	13,1	16,4	13,8	17,1

Tabeli 5 kohaselt diagnoositi südamepuudulikkus meestel ja naistel sarnaselt 55-aastaselt. Jällegi on haridustaseme ja tervisliku seisundi vahel täheldatav seos – kõrgema haridusega isikuid oli rohkem tervete hulgas.

Tabel 5: Südamepuudulikkuse valitud tunnuste osakaalud ja keskmised haigetel ja tervetel geenidoonoritel soo lõikes.

	Mehed		Naised	
	Haiged	Terved	Haiged	Terved
Suitsetajad	68,5%	62,4%	32,0%	36,2%
Alkoholi tarvitajad	94,3%	96,1%	77,6%	88,5%
Põhiharidus	4,8%	1,3%	3,4%	0,7%
Keskharidus	42,6%	36,1%	40,0%	28,1%
Kõrgem haridus	52,6%	62,6%	56,6%	71,2%
Keskmine vanus haigestumisel	55	-	55	-
Keskmine süstoolne vererõhk	138	129	135	122
Keskmine diastoolne vererõhk	84	81	82	77
Keskmine jälgimisperiood	15,0	16,5	17,0	17,1

Tabeli 6 põhjal 2. tüüpi diabeeti haigestuti suhteliselt nooremalt – mehed keskmiselt 52 ja naistel 53-aastaselt. Erinevalt eespool vaadatud haigustest oli jälgimisperiood tervetel mõnevõrra lühem kui haigetel, mis võib olla seotud haiguse varasema avaldumisega. Samuti olid diabeedihaigetel keskmiselt kõrgemad vererõhu näitajad. Madalama haridustasemega geenidoonorite osakaal oli siin eriti silmatorkav just haigete seas.

Tabel 6: 2. tüüpi diabeet valitud tunnuste osakaalud ja keskmised haigetel ja tervetel geenidoonoritel soo lõikes.

	Mehed		Naised	
	Haiged	Terved	Haiged	Terved
Suitsetajad	69,7%	63,3%	33,4%	35,1%
Alkoholi tarvitajad	96,6%	96,0%	80,9%	87,5%
Põhiharidus	2,3%	2,3%	2,5%	1,0%
Keskharidus	41,9%	37,7%	37,2%	30,3%
Kõrgem haridus	55,8%	60,0%	60,4%	68,7%
Keskmine vanus haigestumisel	52	-	53	-
Keskmine süstoolne vererõhk	138	130	134	124
Keskmine diastoolne vererõhk	85	81	83	77
Keskmine jälgimisperiood	16,5	15,8	17,5	16,9

## 10.2 Kardiovaskulaarse haiguse riskitegurid

Naistele koostati Coxi võrdeliste riskide mudel 18-ne tunnusega, lõplikus mudelis osutus statistiliselt oluliseks 12 tunnust - lisaks vanusele ja KVH PRS-le kaheksa metaboliiti ja kaks küsimustiku tunnust. Meeste andmestikust on mudeli sobitamisel kasutatud 15 tunnust ning lõplikku mudelisse jäi 11 tunnust: vanus, koronaararteri PRS, kuus metaboliiti ja kolm küsimustikupõhist näitajat.

Naiste mudeli tulemused on välja toodud tabelis 10, kus tunnus S\_LDL\_PL viitab fosfolipiidide hulgale väikestes LDL-osakestes. Osutub, et naiste puhul on suurimaks riskiteguriks alkoholi tarbimine, mis suurendab kardiovaskulaarse haiguse riski 19,2% võrreldes mittetarbijaga. Suitsetamine on samuti oluline – suitsetajatel on 17,7% kõrgem risk. Kaitsvateks teguriteks on histidiin, atsetaat ja albumiin, mille kõrgemad tasemed on seotud riski vähenemisega vastavalt 3,5%, 11,3% ja 9,0% võrra. Ülejäänud tunnuste puhul jääb riski tõus vahemikku 7,6%-12,3%. Mudeli C-indeksi väärtus tuli 0,769.

Tabel 7: Kardiovaskulaarse haiguse riskitegurid naistel.

<b>Tunnus</b>	$e^{\hat{\beta}}$	<b>95% usaldusintervall</b>
Vanus	1,076	(1,072;1,080)
Suitsetamise staatus	1,177	(1,092; 1,269)
S_LDL_PL	1,082	(1,038; 1,129)
Püroviin	1,096	(1,035; 1,162)
KV haiguse PRS	1,096	(1,059; 1,135)
Histidiin	0,965	(0,932; 0,999)
Glükoproteiini atsetool	1,107	(1,064; 1,153)
Glükoos	1,123	(1,081; 1,168)
Atsetoon	1,082	(1,044; 1,121)
Atsetaat	0,887	(0,852; 0,924)
Alkoholi tarbimine	1,192	(1,090; 1,303)
Albumiin	0,910	(0,877; 0,945)

Tabelis 8 on kujutatud meeste Coxi mudeli tunnuste riskimäärad ja nende 95% usaldusintervallid. Tunnus TOTAL\_BCAA viitab BCAA-de kontsentratsioonile, L\_LDL\_CE\_pct kolesterüülestrite osakaalule kogulipiididest suurtes LDL-osakestes ja S\_HDL\_FC\_pct vaba kolesterooli osakaalule kogulipiididest väikestes HDL-osakestes. Meestel on suurima efektiga riskitegur suitsetamine, mis suurendab kardiovaskulaarse haiguse riski 28,0% võrreldes mittersuitsetajatega. Kaitsvateks teguriteks osutusid albumiin, Total\_BCAA ja L\_LDL\_CE\_pct. Nende suuremine ühe ühiku võrra vähendab riski vastavalt 9,4%, 11,9% ja 10,7%. Ülejäänute tunnuste puhul jääb riski tõus vahemikku 0,7%-15,8%. Mudeli headust kirjeldava C-indeksi väärtus oli 0,768.

Tabel 8: Kardiovaskulaarse haiguse riskitegurid meestel.

<b>Tunnus</b>	$e^{\hat{\beta}}$	<b>95% usaldusintervall</b>
Vanus	1,064	(1,058; 1,069)
KAH PRS	1,109	(1,057; 1,164)
TOTAL_BCAA	0,881	(0,833; 0,932)
Glükoos	1,119	(1,066; 1,174)
Albumiin	0,906	(0,864; 0,951)
Glükoproteiini atsetool	1,159	(1,104; 1,217)
L_LDL_CE_pct	0,893	(0,849; 0,940)
S_HDL_FC_pct	1,134	(1,079; 1,192)
Suitsetamise staatus	1,273	(1,152; 1,407)
Kehamassiindeks	1,021	(1,009; 1,033)
Süstoolne vererõhk	1,007	(1,004; 1,009)

### 10.2.1 Vanusel ja PRS-idel põhinevad mudelid

Edasi on hinnatud lihtsustatud mudel, mis sisaldas vaid vanust ja PRS-e. Sellise mudeli tulemused nii meestele kui ka naistele on esitatud tabelis 9. Tabelis tähistab sidekriip statistiliselt ebaolulisi tunnuseid. Naistel osutus kõige mõjukamaks lisaks vanusele KV haiguse PRS (8,4%), millele järgnesid koronaararteri haiguse PRS (5,6%), südamepuudulikkuse PRS (4,8%) ja 2. tüüpi diabeedi PRS (4,2%). Meestel tuli kõige suurema mõjuga tunnuseks koronaararteri haiguse PRS, tõstes riski 9,3%. Kõige nõrgem mõju meestel oli LDL-kolesterooli PRS-il, mis ühe ühiku suurenemisel tõustab riski 5,0%. Mudelite C-indeksid olid 0,758 (naistel) ja 0,745 (meestel).

Tabel 9: Kardiovaskulaarse haiguse polügeenseid riskiskoore ning vanust sisaldava mudeli tulemused naistel ja meestel.

Tunnus	Sugu	$e^{\hat{\beta}}$ ja 95% usaldusintervall
Vanus	Mehed	1,075 (1,070; 1,080)
	Naised	1,084 (1,080; 1,087)
KVH PRS	Mehed	1,075 (1,025; 1,127)
	Naised	1,084 (1,044; 1,124)
2. tüüpi diabeedi PRS	Mehed	1,073 (1,023; 1,126)
	Naised	1,042 (1,006; 1,079)
Südamepuudulikkuse PRS	Mehed	-
	Naised	1,048 (1,010; 1,087)
KAH PRS	Mehed	1,093 (1,040; 1,149)
	Naised	1,056 (1,019; 1,095)
LDL-kolesterooli PRS	Mehed	1,050 (1,002; 1,099)
	Naised	-

### 10.3 Kodade virvendusarütmia riskitegurid

Naistel on esmane Coxi võrdeliste riskide mudel sobitatud 14 tunnusega ja kõik need tunnused jäid lõplikku mudelisse. Oluliste tunnuste seas oli vanus, kaks PRS-i, kuus metaboliiti ja viis küsimustiku tunnust. Meeste puhul kasutati 16 tunnust ja lõplik mudel koosnes 12 tunnusest - vanus, üks PRS, seitse metaboliiti ja kolm küsimustiku tunnust.

Naiste mudeli tulemused on toodud tabelis 10. Tunnus XS\_VLDL\_PL\_pct viitab fosfolipiidide osakaalule väga väikestes väga väikse tihedusega lipoteiinide (*Very low density lipoprotein*, VLDL) osakestes. Selgub, et suurimaks riskiteguriks on alkoholi tarbimine, suurendades riski 27,4% võrreldes alkoholi mittetarbijaga. Efekti suuruse poolest järgmine tunnus oli suitsetamine, suurendades haigestumise riski 27,0%. KVA ja koronaararteri haiguse PRS-ide ühe standardvea võrra suurenemine toob kaasa haiguse riski suurenemise vastavalt 16,1% ja 9,3%. Ülejäänute tunnuste

puhul jääb riskitõus vahemikku 0,5%-10,6%.

Riski vähendavad tegurid on leutsiin, albumiin ja kõrgeim omandatud haridus. Leutsiini ja albumiini suurem tase vähendab riski vastavalt 11,1% ja 16,2%. Kõrgeima omandatud hariduse puhul, kus baastase on põhiharidus, on oluliselt väiksem risk indiviidil, kes on omandanud kõrgema haridustaseme kui keskharidus. Võrreldes omavahel põhi- ja keskharidusega naisi, ei olnud mõju statistiliselt oluline. Naiste mudeli C-indeks oli 0,833, mis viitab ülimalt heale diskrimineerimisvõimele.

Tabel 10: Kodade virvendusarütmia riskitegurid naistel.

<b>Tunnus</b>	$e^{\hat{\beta}}$	<b>95% usaldusintervall</b>
Vanus	1,106	(1,100; 1,112)
KVA PRS	1,161	(1,111; 1,212)
KAH PRS	1,093	(1,046; 1,143)
Glükoos	1,094	(1,042; 1,148)
Leutsiin	0,889	(0,847; 0,933)
Atsetoon	1,096	(1,048; 1,148)
Albumiin	0,838	(0,797; 0,881)
Glükoproteiini atsetool	1,103	(1,044; 1,165)
XS_VLDL_PL_pct	1,062	(1,012; 1,114)
Suitsetamise staatus	1,270	(1,148; 1,406)
Alkoholi tarbimine	1,274	(1,139; 1,425)
Kehamassiindeks	1,031	(1,022; 1,040)
Süstoolne vererõhk	1,005	(1,003; 1,008)
Haridustase: kõrgem haridus	0,778	(0,620; 0,975)
Haridustase: keskharidus	0,899	(0,718; 1,126)

Meeste mudeli tulemused on tabelis 11, milles tunnus L\_LDL\_CE\_pct näitab kolesteriülestrite osakaalu kogulipiididest suurtes LDL-osakestes ja S\_LDL\_PL\_pct fosfolipiidide osakaalu kogulipiididest väikestes LDL-osakestes. Kõige enam suurendab riski meeste seas suitsetamine, mis tõstab haigestumise riski 32,1%. Teiste riski tõstvate näitajate korral jääb riskitõus vahemikku 0,8%-15,7%. Riski vähendavad aga leutsiini, albumiini ja L\_LDL\_CE\_pct suurenemine vastavalt 9,1%, 9,6% ja 11,2%. Mudeli headust kirjeldava C-indeksi väärtus oli 0,81.

Tabel 11: Kodade virvendusarütmia riskitegurid meestel.

<b>Tunnus</b>	$e^{\hat{\beta}}$	<b>95% usaldusintervall</b>
Vanus	1,083	(1,077; 1,090)
KAH PRS	1,075	(1,020; 1,132)
Leutsiin	0,909	(0,856; 0,965)
Glükoos	1,145	(1,085; 1,208)
$\beta$ -hüdrosübutüraat	1,127	(1,057; 1,202)
Albumiin	0,904	(0,854; 0,958)
Glükoproteiini atsetool	1,113	(1,054; 1,176)
L_LDL_CE_pct	0,888	(0,837; 0,942)
S_LDL_PL_pct	1,157	(1,086; 1,233)
Suitsetamise staatus	1,321	(1,180; 1,479)
Kehamassiindeks	1,038	(1,026; 1,050)
Süstoolne vererõhk	1,008	(1,005; 1,011)

### 10.3.1 Vanusel ja PRS-idel põhinevad mudelid

Kodade virvendusarütmia vanust ja PRS-e sisaldavate mudelite tulemused on tabelis 15, kus sidekriipsud tähistavad statistiliselt ebaolulisi tunnuseid. Aastane vanuse kasv suurendab naistel riski 11,8%. PRS-idest osutus kõige mõjukamaks KVA PRS, suurendades riski 16,1%. Koronaararteri haiguse ja kehamassiindeksi PRS-id suu-

rendavad haigestumise riski vastavalt 9,7% ja 13,2%. Naiste mudeli C-indeks tuli 0,82.

Meeste mudelis osutusid oluliseks vanus, 2. tüüpi diabeedi PRS, südamepuudulikkuse PRS, koronaararteri haiguse PRS, kehamassiindeksi PRS, kus standardhälbe ühiku suurenemine tõstab riski vastavalt 9,7%, 9,2%, 6,9%, 6,8%, 7,3%. Ka triglütseriidide PRS osutus oluliseks, kuid selle puhul väheneb risk standardhälbe ühiku suurenemisel 6,2%. Meeste mudeli C-indeksi väärtus oli 0,78.

Tabel 12: Kodade virvendusarütmia polügeenseid riskiskoore ning vanust sisaldava mudeli tulemused naistel ja meestel.

Tunnus	Sugu	$e^{\hat{\beta}}$ ja 95% usaldusintervall
Vanus	Mehed	1,097 (1,091; 1,103)
	Naised	1,118 (1,113; 1,124)
KVA PRS	Mehed	-
	Naised	1,161 (1,112; 1,212)
KAH PRS	Mehed	1,068 (1,014; 1,126)
	Naised	1,097 (1,049; 1,146)
Kehamassiindeksi PRS	Mehed	1,073 (1,019; 1,30)
	Naised	1,131 (1,082; 1,183)
2. tüüpi diabeedi PRS	Mehed	1,092 (1,034; 1,510)
	Naised	-
Südamepuudulikkuse PRS	Mehed	1,069 (1,015; 1,126)
	Naised	-
Triglütseriidide PRS	Mehed	0,938 (0,890; 0,988)
	Naised	-

## 10.4 Koronaararteri haiguse riskitegurid

Naistel kasutati LASSO-st tulenevalt esimese mudeli loomisel 20 tunnust. Coxi võrdeliste riskide lõplikku mudelisse jäi 15 - vanus, kaks PRS-i, üheksa metaboliiti

ja kolm küsimustiku tunnust. Meestel on esmane Coxi mudel sobitatud 21 tunnusega, neist olulisteks jäi 16. Meestel oli tunnuste seas vanus, kaks PRS-i, üheksa metaboliiti ja neli küsimustiku tunnust.

Naiste mudeli tulemused on tabelis 13. Osutub, et suurima mõjuga riskiteguriks naiste seas on suitsetamine, tõstes riski 65,3% võrreldes mittesuitsetajaga. Metaboliitidest tõstab kõige rohkem riski glükoos, täpsemalt 20,2%. Ülejäänute riski tõstvate tunnuste puhul jääb riskitõus vahemikku 1,1%-19,7%.

Riski vähendavad kõrgeim omandatud haridustase, leutsiin, albumiin ja oomega-3 rasvhappe osakaal kogu rasvhapetest (tunnus Omega\_3\_pct). Kõrgeima omandatud hariduse puhul on baastase põhikooli haridus, seega saab öelda, et naisel, kelle kõrgeim omandatud haridus on kõrgem kui keskharidus, on 43,3% väiksem risk haigestumiseks, kui põhiharidusega naisel. Sarnaselt eelnevale on naisel, kellel on kõrgeim omandatud haridus keskharidus, 25,8% väiksem risk võrreldes põhihariduse naistega. Suurenenud leutsiini, albumiini ja oomega-3 rasvhappe osakaal kogu rasvhapetest vähendavad riski vastavalt 10,3%, 26,1% ja 9,3%. Mudeli diskrimineerimisvõime oli väga hea, C-indeks = 0,868.

Tabel 13: Koronaararteri haiguse riskitegurid naistel.

<b>Tunnus</b>	$e^{\hat{\beta}}$	<b>95% usaldusintervall</b>
Vanus	1,111	(1,101; 1,120)
KAH PRS	1,197	(1,121; 1,278)
Hüpertensiooni PRS	1,128	(1,057; 1,203)
Omega_3_pct	0,907	(0,851; 0,967)
Leutsiin	0,897	(0,838; 0,961)
Glükoos	1,202	(1,123; 1,288)
Piimhape	1,093	(1,024; 1,166)
$\beta$ -hüdrosübutüraat	1,150	(1,070; 1,235)
Albumiin	0,739	(0,689; 0,792)
Glükoproteiini atsetool	1,165	(1,080; 1,258)
XL_VLDL_PL_pct	1,137	(1,050; 1,231)
XS_VLDL_PL_pct	1,133	(1,050; 1,219)
Suitsetamise staatus	1,653	(1,438; 1,899)
Süstoolne vererõhk	1,011	(1,007; 1,014)
Haridustase: kõrgem haridus	0,567	(0,430; 0,747)
Haridustase: keskharidus	0,742	(0,565; 0,975)

Meeste tulemused on tabelis 14, kus tunnus S\_HDL\_FC\_pct näitab vaba kolesterooli osakaalule kogulipiididest väikestes HDL-osakestes, S\_VLDL\_CE\_pct vaba kolesterooli osakaalule väikestes VLDL-osakestes ja XS\_VLDL\_PL\_pct fosfolipiidide osakaalule väga väikestes VLDL-osakestes. Taaskord on meestel suurima mõjuga riskitegur suitsetamine, suurendades haigestumise riski 38,3% võrra. Sellele järgneb mõju suuruse poolest koronaararteri haiguse PRS, suurendades haiguse riski 22,9% ühe standardhälbe ühiku tõustes. Samuti on oluline roll KV haiguse PRS-il, kuid selle mõju on väiksem, tõstes riski 6,9%. Metaboliitidest on olulisel kohal järgmised näitajad:  $\beta$ -hüdrosübutüraat, S\_HDL\_FC\_pct, glükoos, S\_VLDL\_CE\_pct, glükoproteiini atsetool ja

XS\_VLDL\_PL\_pct. Eelnevate metaboliitide korral jääb riskitõus vahemikku 10,9%-19,3%. Ühtlasi tõuseb risk ühiku suurenemisel vanuse (7,7%), kehamassiindeksi (2,3%) ja süstoolse vererõhu (0,8%) puhul.

Riski vähendavad tunnused on meestel naistega samad. Ühtlasi leidub oluline seos haigestumise ja hariduse vahel: võrreldes baastasemega (s.t põhiharidusega), on kõrgharitud mehel 32,5% väiksem risk haigestuda. Küll aga ei leitud olulist seost põhi- ja keskkharidusega meeste vahel. Metaboliitidest vähendavad riski albumiin, leutsiin ja oomega-3 rasvhappe osakaal. Nende tasemete suurenemine vähendab riski vastavalt 16,6%, 13,5% ja 8,0%. Mudeli C-indeks oli 0,829.

Tabel 14: Koronaararteri haiguse riskitegurid meestel.

<b>Tunnus</b>	$e^{\hat{\beta}}$	<b>95% usaldusintervall</b>
Vanus	1,077	(1,070; 1,085)
KAH PRS	1,229	(1,152; 1,310)
KVH PRS	1,069	(1,003; 1,140)
Omega_3_pct	0,920	(0,866; 0,977)
Leutsiin	0,865	(0,866; 0,929)
Glükoos	1,147	(1,077; 1,223)
$\beta$ -hüdroksübutüraat	1,193	(1,107; 1,286)
Albumiin	0,834	(0,782; 0,890)
Glükoproteiini atsetool	1,131	(1,053; 1,215)
S_VLDL_CE_pct	1,131	(1,057; 1,210)
XS_VLDL_PL_pct	1,109	(1,040; 1,183)
S_HDL_FC_pct	1,157	(1,078; 1,243)
Suitsetamise staatus	1,382	(1,207; 1,582)
Kehamassiindeks	1,023	(1,207; 1,038)
Süstoolne vererõhk	1,008	(1,004; 1,011)
Haridustase: kõrgem haridus	0,676	(0,516; 0,884)
Haridustase: keskkharidus	0,770	(0,589; 1,006)

### 10.4.1 Vanusel ja PRS-idel põhinevad mudelid

Tabelis 15 on koronaararteri haiguse vanust ja PRS-e sisaldavate mudelite tulemused meestel ja naistel. Naistel jäi lõplikku mudelisse vanus, koronaararteri haiguse PRS, hüpertensiooni PRS ja kehamassiindeksi PRS, kusjuures PRS-ide ühe ühiku standardhälbe suurenemine tõstab haigestumise riski vastavalt 23,0%, 17,8% ja 8,0%. Naistel tuli mudeli C-indeks 0,83.

Meeste mudelisse jäi oluliste tunnustena vanus, KVA PRS, südamepuudulikkuse PRS, koronaararteri haiguse PRS ja hüpertensiooni PRS. Selgus, et viimase kolme PRS-i suurenemine ühe standardhälbe võrra suurendab riski vastavalt 9,1%, 25,3% ja 9,9%. KVA PRS-i suurenemine hoopiski vähendab riski 7,0%. Meeste mudeli C-indeksi väärtuseks tuli 0,79.

Tabel 15: Koronaaarteri haiguse naiste ja meeste polügeenseid riskiskoore ning vanust sisaldava mudeli tulemused.

Tunnus	Sugu	$e^{\hat{\beta}}$ ja 95% usaldusintervall
Vanus	Mehed	1,096 (1,090; 1,103)
	Naised	1,130 (1,122; 1,138)
KAH PRS	Mehed	1,253 (1,177; 1,333)
	Naised	1,230 (1,153; 1,312)
Kehamassiindeksi PRS	Mehed	-
	Naised	1,080 (1,014; 1,151)
Hüpertensiooni PRS	Mehed	1,099 (1,030; 1,173)
	Naised	1,178 (1,105; 1,256)
Südamepuudulikkuse PRS	Mehed	1,091 (1,023; 1,164)
	Naised	-
KVA PRS	Mehed	0,930 (0,874; 0,990)
	Naised	-

## 10.5 Südamepuudulikkuse riskitegurid

Naiste Coxi võrdeliste riskide mudeli koostamisel on kasutatud 18 tunnust, nendest jäi lõplikku mudelisse 14. Meestel on esmane mudel sobitatud 24 tunnusega ja neist 15 jäi lõplikku mudelisse statistiliselt olulistena.

Naiste Coxi võrdeliste riskide mudeli tulemused on välja toodud tabelis 16, kus tunnus L\_LDL\_TG\_pct viitab triglütseriidide osakaalule kogulipiididest suurtes LDL-osakestes. Siingi on naistel suurimate efektidega riskitegurid suitsetamine ja alkoholitarbimine: suitsetaval naisel on 30,8% suurem risk kui mittesuitsetajal ja alkoholi tarvitajal on risk 15,7% suurem kui alkoholi mittetarbival naisel. PRS-idest osutusid olulisteks koronaararteri haiguse ja hüpertensiooni PRS-id, kus mõlema näitaja ühe ühiku standardhälbe võrra suurenemine tõstab riski vastavalt 5,2% ja 10,6%. Metaboliitidest on glükoos, atsetoon, glükoproteiini atsetool ja L\_LDL\_TG\_pct riski suurendavad näitajad ning nende suurenemine tõstab südamepuudulikkuse tekkeriski vastavalt 13,1%, 8,2%, 13,0% ja 4,0%. Viimaste riskiteguritena saab välja tuua vanuse (7,7%), kehamassiindeksi (3,9%) ja süstoolse vererõhu (0,8%).

Naiste seas vähendavad riski kolm näitajat. Kõige suurema mõjuga on albumiin, mille korral risk väheneb 10,9%. Ülejäänud kahe tunnuse puhul iga ühiku suurenemine vähendab haigestumise riski histidiini korral 5,4% ja oomega-3 rasvhappe osakaalu puhul 4,3%.

Tabel 16: Südamepuudulikkuse riskitegurid naistel.

<b>Tunnus</b>	$e^{\hat{\beta}}$	<b>95% usaldusintervall</b>
Vanus	1,077	(1,073; 1,082)
KAH PRS	1,052	(1,015; 1,090)
Hüpertensiooni PRS	1,106	(1,067; 1,147)
Omega_3_pct	0,957	(0,923; 0,992)
Histidiin	0,946	(0,913; 0,980)
Glükoos	1,131	(1,088; 1,175)
Albumiin	0,892	(0,857; 0,928)
Atsetoon	1,082	(1,044; 1,122)
Glükoproteiini atsetool	1,130	(1,083; 1,179)
L_LDL_TG_pct	1,040	(1,002; 1,079)
Suitsetamise staatus	1,308	(1,210; 1,415)
Alkoholi tarvitamine	1,157	(1,061; 1,260)
Kehamassiindeks	1,039	(1,032; 1,046)
Süstoolne vererõhk	1,008	(1,006; 1,010)

Meeste Coxi võrdeliste riskide mudeli tulemused on esitatud tabelis 17, kus tunnus IDL\_CE\_pct näitab kolesterüülestrite osakaalu keskmise tihedusega lipoproteiinide (*Intermediate density lipoprotein*, IDL) osakestes, L\_LDL\_CE\_pct kolesterüülestrite osakaalu suurtes LDL-osakestes, S\_LDL\_PL\_pct fosfolipiidide osakaalu väikestes LDL-osakestes ja S\_HDL\_FC\_pct vaba kolesterooli osakaalul väikestes HDL-osakestes. Kõige enam tõstab haigestumise riski meeste seas suitsetamine: isikul, kes suitsetab on 35,2% suurem risk haigestuda. Metaboliitidest suurendavad riski järgmised näitajad: glükoproteiini atsetool, glükoos, S\_LDL\_PL\_pct, atsetoon ja S\_HDL\_FC\_pct, mille korral risk suureneb vastavalt 17,5%, 14,9%, 12,5%, 11,3% ja 7,4%. Ühtlasi suureneb risk 7,5% vanusega iga aasta, kehamassiindeksi tõus tõstab riski 5,2% ja risk tõuseb 1,1%, kui süstoolse vererõhu näit tõuseb ühe ühiku võrra.

Meeste seas leidub viis tunnust, mille puhul südamepuudulikkuse risk väheneb.

Need viis tunnust on kõik metaboliidid. Leutsiin, albumiin, Omega\_3\_pct, IDL\_CE\_pct ja L\_LDL\_CE\_pct vähendavad südamepuudulikkuse tekkeriski vastavalt 14,1%, 8,2%, 8,4%, 7,5% ja 5,5%.

Tabel 17: Südamepuudulikkuse riskitegurid meestel.

<b>Tunnus</b>	$e^{\hat{\beta}}$	<b>95% usaldusintervall</b>
Vanus	1,075	(1,070; 1,081)
KAH PRS	1,109	(1,057; 1,163)
Omega_3_pct	0,916	(0,873; 0,961)
Leutsiin	0,859	(0,813; 0,908)
Glükoos	1,149	(1,095; 1,205)
Atsetoon	1,113	(1,063; 0,165)
Albumiin	0,918	(0,870; 0,970)
Glükoproteiini atsetool	1,174	(1,112; 1,240)
IDL_CE_pct	0,925	(0,881; 0,972)
L_LDL_CE_pct	0,945	(0,895; 0,998)
S_LDL_PL_pct	1,125	(1,059; 1,194)
S_HDL_FC_pct	1,074	(1,018; 1,134)
Suitsetamise staatus	1,352	(1,221; 1,134)
Kehamassiindeks	1,052	(1,040; 1,063)
Süstoolne vererõhk	1,011	(1,008; 1,014)

### 10.5.1 Vanusel ja PRS-idel põhinevad mudelid

Südamepuudulikkuse vanust ja PRS-e sisaldavate mudelite tulemused on toodud tabelis 18. Meestel ja naistel osutusid olulisteks täpselt samad tunnused (vanus, KAH PRS, kehamassiindeksi PRS, hüpertensiooni PRS, HDL-kolesterooli PRS ja südamepuudulikkuse PRS). PRS-ide tõus naistel suurendab riski vastavalt 5,8%, 9,8%, 14,6%, 4,1% ja 5,6%. Meestel tõstab riski koronaararteri haiguse PRS 10,3%, kehamassiindeksi PRS 10,5%, hüpertensiooni PRS 6,0%, südamepuudulikkuse PRS 6,8% ja HDL-kolesterooli PRS vähendab riski 4,7% võrra. Naistel ja meestel tuli C-indeks samuti sama, 0,78.

Tabel 18: Südamepuudulikkuse naiste ja meeste polügeenseid riskiskoore ning vanust sisaldava mudeli tulemused.

Tunnus	Sugu	$e^{\hat{\beta}}$ ja 95% usaldusintervall
Vanus	Mehed	1,089 (1,084; 1,094)
	Naised	1,092 (1,088; 1,096)
KAH PRS	Mehed	1,103 (1,051; 1,157)
	Naised	1,058 (1,021; 1,096)
Kehamassiindeksi PRS	Mehed	1,105 (1,054; 1,158)
	Naised	1,098 (1,059; 1,137)
Hüpertensiooni PRS	Mehed	1,060 (1,009; 1,113)
	Naised	1,146 (1,104; 1,189)
HDL-kolesterooli PRS	Mehed	0,953 (0,909; 0,998)
	Naised	1,041 (1,004; 1,078)
Südamepuudulikkuse PRS	Mehed	1,068 (1,017; 1,121)
	Naised	1,056 (1,017; 1,096)

## 10.6 2. tüüpi diabeedi riskitegurid

Naistel ja meestel on esmane Coxi võrdeliste riskide mudel koostatud 20 muutujaga, naistel jäi 17 olulisteks ja meestel 16. Naiste lõplikus mudelis on vanus, kaks PRS-i, kümme metaboliiti ja neli küsimustiku tunnust. Meestel on vanus, kolm PRS-i, üheksa metaboliiti ja kolm küsimustikust saadud näitajat.

Tabelis 19 on toodud naiste lõplikku mudelisse kuuluvad tunnused, nende riskimäärad ja 95% usaldusintervallid, kus tunnus M\_HDL\_P on keskmise suurusega HDL-osakeste kontsentratsioon, XXL\_VLDL\_FC\_pct vaba kolesterooli osakaal kogulipiididest väga suurtes VLDL-osakestes, XS\_VLDL\_CE\_pct kolesterüülest-rite osakaal kogulipiididest väga väikestes VLDL-osakestes ja S\_HDL\_PL\_pct fosfolipiidide osakaal väikestes HDL-osakestes. Südamehaigustel oli tavaliselt nais- te seas suurimat riski põhjustav tegur suitsetamine või alkoholi tarbimine, kuid 2. tüüpi diabeedi puhul on kõrgel glükoosi tasemel kõige suurem risk: kahe nai- se korral, kellest ühel on glükoosi tase ühe ühiku võrra suurem ja teised näitajad samad, on 2. tüüpi diabeeti haigestumise risk 59,4% suurem. Suitsetamine ja al- koholi tarbimine on siiski olulised tunnused, tõstes riski vastavalt 12,9% ja 27,6% võrreldes mittesuitsetaja ja alkoholi mittetarbiva naisega. Glükeeritud hemoglobii- ni ja 2. tüüpi diabeedi PRS-idel on oluline roll riski hindamisel: nende üheühikuline tõus toob kaasa riski suurenemise vastavalt 11,5% ja 28,4%. Peale vanuse, keha- massiindeksi ja diastoolse vererõhu, mis tõstavad riski 3,7%, 7,0% ja 0,5%, on ka olulised piimhappe, atsetooni ja glükoproteiini atsetooli näitajad, põhjustades riski suurenemise vastavalt 9,8%, 11,7%, 21,4%.

Naistel on 2. tüüpi diabeeti haigestumise riski vähendavad näitajad atsetaat, XS\_VLDL\_FC\_pct, S\_HDL\_PL\_pct, glükotsiin, XXL\_VLDL\_FC\_pct ja M\_HDL\_P. Eelnevad on kõik metaboliidid ja juhul, kui indiviididel on kõik teised näitajad samad, kuid üks eelnevatest on ühe ühiku võrra kõrgem, siis on kõrgema väärtusega naisel haigestumise risk 17,2%, 16,3%, 14,4%, 10,4%, 8,1% ning 7,8% väiksem.

Tabel 19: 2. tüüpi diabeedi riskitegurid naistel.

<b>Tunnus</b>	$e^{\hat{\beta}}$	<b>95% usaldusintervall</b>
Vanus	1,037	(1,032; 1,042)
2. tüüpi diabeedi PRS	1,284	(1,227; 1,344)
Glükohemoglobiini PRS	1,115	(1,066; 1,167)
Glükotsiin	0,896	(0,856; 0,937)
Glükoos	1,594	(1,513; 1,680)
Piimhape	1,098	(1,046; 1,153)
Atsetaat	0,828	(0,783; 0,875)
Atsetoon	1,117	(1,066; 1,172)
Glükoproteiini atsetool	1,214	(1,139; 1,293)
M_HDL_P	0,922	(0,866; 0,982)
XXL_VLDL_FC_pct	0,919	(0,873; 0,967)
XS_VLDL_CE_pct	0,837	(0,785; 0,893)
S_HDL_PL_pct	0,856	(0,803; 0,913)
Suitsetamise staatus	1,129	(1,024; 1,246)
Alkoholi tarvitamine	1,276	(1,137; 1,432)
Kehamassiindeks	1,070	(1,062; 1,079)
Diastoolne vererõhk	1,005	(1,001; 1,009)

Meeste lõpliku mudeli tulemused on tabelis 20, kus M\_HDL\_P taaskord keskmise suurusega HDL-osakeste kontsentratsioon, IDL\_CE\_pct kolesterüülestrite osakaal IDL-osakestes ja S\_HDL\_FC\_pct vaba kolesterooli osakaal kogulipiididest väikestes HDL-osakestes. Meestelgi põhjustab glükoosi taseme ühikuline tõus 59,7% suurema riski, kui teised näitajad jäävad samaks. Kolmest PRS-ist on suurima mõjuga just 2. tüüpi diabeedi PRS, mis toob kaasa 30,2% suurema riski. Glükeeritud hemoglobiini ja koronaararteri haiguse PRS-ide suurenemine tõstab riski vastavalt 10,6% ja 6,9%. Võrreldes kahte meest, kellest üks suitseb ja teine ei suitseta, on suitsetajal 36,5% suurem 2. tüüpi diabeedi tekkerisk. Metaboliitidest on olulised riski

suurendajad  $\beta$ -hüdrosübutüraat, glükeeritud atsetoon ja S\_HDL\_FC\_pct. Kui kõigi teiste näitajate väärtused on samad ja tõusnud on ühiku võrra vaid eelnevalt loetletud näitajad, siis risk on suurem vastavalt 17,1%, 14,7% ja 9,8%. Muidugi on mõju ka vanusel, kehamassiindeksil ja diastoolsel vererõhul. Nende tõus suurendab riski vastavalt 4,4%, 10,0% ja 1,0%.

Meestel vähendavad 2.tüüpi diabeeti haigestumist järgnevad metaboolmika näitajad: küllastumata rasvhapete aste, histidiin, atsetaat, M\_HDL\_P ja IDL\_CE\_pct. Kui kõigi teiste mudelis olevate tunnuste väärtused on samad ja loetelus olevate tunnuste väärtus on tõusunud ühiku võrra, siis on risk väiksem vastavalt 6,7%, 9,2%, 12,2%, 11,0%, 7,3%.

Tabel 20: 2. tüüpi diabeedi riskitegurid meestel.

<b>Tunnus</b>	$e^{\hat{\beta}}$	<b>95% usaldusintervall</b>
Vanus	1,044	(1,037; 1,051)
2. tüüpi diabeedi PRS	1,302	(1,217; 1,393)
KAH PRS	1,069	(1,001; 1,142)
Glükohemoglobiini PRS	1,101	(1,034; 1,172)
Küllastumata rasvhapete aste	0,933	(0,877; 0,993)
Histidiin	0,908	(0,853; 0,967)
Glükoos	1,597	(1,487; 1,716)
$\beta$ -hüdrosübutüraat	1,171	(1,075; 1,276)
Atsetaat	0,878	(0,820; 0,939)
Glükoproteiini atsetool	1,147	(1,062; 1,239)
M_HDL_P	0,891	(0,833; 0,952)
IDL_CE_pct	0,927	(0,865; 0,994)
S_HDL_FC_pct	1,098	(1,021; 1,179)
Suitsetamise staatus	1,365	(1,187; 1,567)
Kehamassiindeks	1,101	(1,087; 1,114)
Diastoolne vererõhk	1,010	(1,004; 1,016)

### 10.6.1 Vanusel ja PRS-idel põhinevad mudelid

Tabelis 21 on välja toodud 2. tüüpi diabeedi naiste ja meeste mudelite tulemused, milles kasutatud on vanust ja PRS-e. Naistel osutus oluliseks seitse tunnust, vanus ja kuus PRS-i: 2. tüüpi diabeedi, südamepuudulikkuse, hüpertensiooni, kehamassiindeksi, trüglütseriidide ja glükeeritud hemoglobiini, mis tõstavad riski vastavalt 36,3%, 5,3%, 6,3%, 16,1%, 8,2% ja 10,8% ning seda iga ühiku standardhälbe tõusu korral. Mudeli põhjal tõuseb iga aasta vanusega risk 6,2%. Naiste mudeli C-indeksi väärtus tuli 0,74.

Meestel tuli oluliseks vanus, 2. tüüpi diabeedi PRS, kehamassiindeksi PRS, glükeeritud hemoglobiini PRS, KVH PRS ja HDL-kolesterooli PRS, mis tõstavad riski vastavalt 5,7%, 35,8%, 17,0%, 12,3% ja 10,5%. Riski vähendab HDL-kolesterooli PRS 6,4% võrra. Meeste mudeli C-indeksi väärtuseks tuli 0,72.

## 10.7 SCORE2 komponentidega mudelid

Viimaks on iga haiguse jaoks nii meestele kui ka naistele hinnatud Coxi võrdeliste riskide mudelid, milles on kasutatud SCORE2 tunnuseid. Nende mudelite C-indekseid võrreldi eespool välja töötatud mudelitega.

Kardiovaskulaarse haiguse korral tulid nii meestel kui ka naistel kõik SCORE2 tunnused olulised. Naiste mudeli C-indeks tuli 0,76 ja meestel 0,75.

Kodade virvendusarütmia mudelis tulid naistel kõik tunnused statistiliselt olulised, kuid meestel osutus HDL-kolesterool ebaoluliseks (p-väärtus 0,23). Naiste ja meeste mudelite C-indeksid olid vastavalt 0,82 ja 0,79.

Koronaararteri haiguse puhul tulid statistiliselt olulised tunnused kõik peale üldkolesterooli ja seda nii naistel kui meestel. Mudeli C-indeksi tuli naistel 0,84 ja meestel 0,79.

Südamepuudulikkuse korral osutus ebaoluliseks üldkolesterool ja seda nii naistel

Tabel 21: 2. tüüpi diabeedi naiste ja meeste polügeenseid riskiskoores ning vanust sisaldavate mudelite tulemused.

<b>Tunnus</b>	<b>Sugu</b>	$e^{\hat{\beta}}$ ja 95% usaldusintervall
Vanus	Mehed	1,057 (1,051; 1,063)
	Naised	1,062 (1,058; 1,066)
2. tüüpi diabeedi PRS	Mehed	1,358 (1,270; 1,451)
	Naised	1,363 (1,300; 1,428)
Südamepuudulikkuse PRS	Mehed	-
	Naised	1,053 (1,004; 1,105)
Hüpertensiooni PRS	Mehed	-
	Naised	1,063 (1,014; 1,113)
Kehamassiindeksi PRS	Mehed	1,170 (1,097; 1,247)
	Naised	1,161 (1,109; 1,215)
Triglütseriidide PRS	Mehed	-
	Naised	1,082 (1,034; 1,132)
Glükeeritud hemoglobiini PRS	Mehed	1,123 (1,054; 1,196)
	Naised	1,108 (1,059; 1,159)
KVH PRS	Mehed	1,105 (1,037; 1,177)
	Naised	-
HDL-kolesterooli PRS	Mehed	0,936 (0,879; 0,998)
	Naised	-

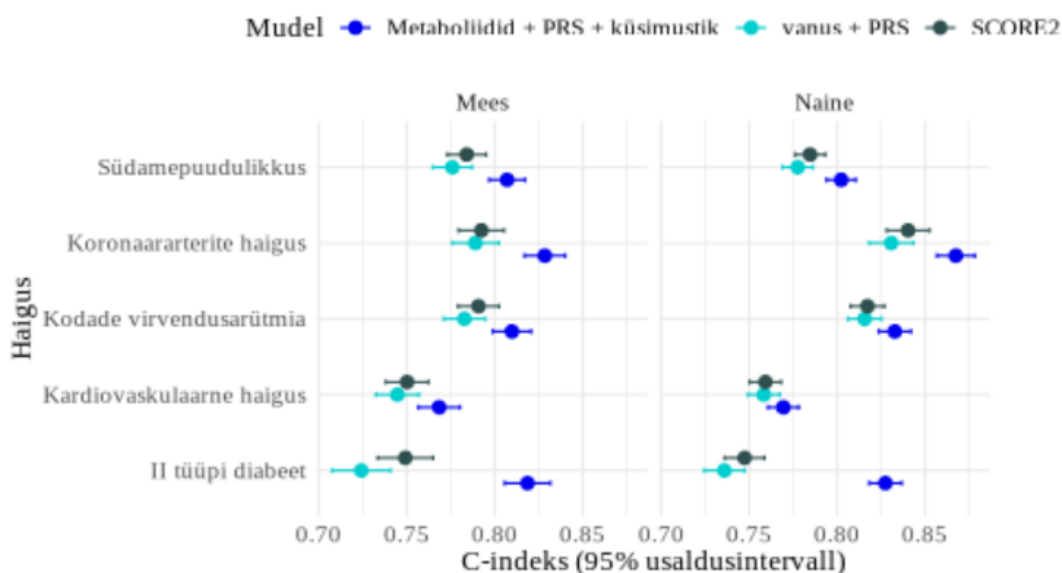
(p-väärtus 0,291) kui meestel (p-väärtus 0,213). Naiste mudeli C-indeks oli 0,76 ja meestel 0,78.

2. tüüpi diabeedi korral tulid kõik SCORE2 komponendid nii naistel kui meestel statistiliselt olulised ja mõlema mudeli C-indeksi oli 0,75.

## 10.8 Mudelite võrdlus

Mudelite sooritusvõimed võtab kokku joonis 1, kus on esitatud C-indeksid ja 95% usaldusintervallid uuritud haiguste ja soo lõikes. Tulemused näitavad, et nii meestel kui naistel saavutab igal korral kõrgema C-indeksi väärtuse mudel, milles on otsitud riskifaktoreid kõigest kolmest andmestikust (küsimustik, PRS-id ja metaboliidid). Eriti hästi tuleb see erinevus välja 2. tüüpi diabeedi puhul, kus meestel on C-indeks 0,82 ja naistel 0,83. Samal ajal kui teiste 2. tüüpi diabeedi mudelite C-indeksi väärtused jäävad nii meestel kui naistel vahemikku 0,73-0,75. Ootuspärane on aga see, et 2. tüüpi diabeedi puhul SCORE2 mudeli C-indeks on madalam kui teistel haigustel, sest see on siiski mõeldud KV haiguse riskihindamiseks. Meeste ja naiste mudelite võrdlemisel on peaaegu kõigil juhtudel meeste mudelite prognoosivõime pisut madalam.

Ühtlasi on näha, et vanust ja PRS-e sisaldavate mudelite ennustusvõimed on küllaltki sarnased SCORE2 tunnustega iga haiguse puhul nii meestel kui naistel. Isegi kui erinevus nende mudelite vahel on väike, siis SCORE2 komponentidega mudeli C-indeksid on suuremad.



Joonis 1: Mudelite C-indeksid ja 95% usaldusintervallid haiguste ja soo lõikes.

Aastal 2024 valmis uuring, milles sooviti näha, kas südamepuudulikkuse riskimudeli sooritusvõime paraneb, kui sellesse lisada metabooloomika näitajad. Töö tulemusena valmis mudel, mille C-indeks oli 0,768. [33] Käesoleva töö tulemustena saadi mudelid südamepuudulikkusele, mille C-indeks oli nii meestel kui ka naistel suurem kui 0,8.

Näitena saab tuua veel ühe 2024. aastal valminud uuringu tulemuse, milles kasutati metaboliite ja SCORE2 mudeli tunnuseid, et prognoosida KV haigust. Sellise lähenemisviisiga saadi mudel, mille C-indeks oli 0,71. [34] Seeegi jääb alla käesolevas töös loodud mudelitele, milles KV haiguse mudelite C-indeksid olid ligikaudu 0,77 nii naistel kui ka meestel.

## Kokkuvõte

Käesoleva bakalaureusetöö eesmärk oli välja töötada nelja kardiovaskulaarse haiguse ja 2. tüüpi diabeedi riski prognoosivad soospetsiifilised mudelid ja nende mudelite tulemusi võrrelda laialdaselt kasutatava kliinilise riskiskoori komponentidega. Mudelite välja töötamisel uuriti, kas ja kui palju paraneb prognoosivõime, kui mudelis on kasutatud metaboolmilisi markereid või polügeenseid riskiskoores. Haigused, millele mudelid loodi on järgnevalt välja toodud: kardiovaskulaarne haigus, kodade virvendusarütmia, koronaararteri haigus, südamepuudulikkus ja 2. tüüpi diabeet. Töös kasutati Tartu Ülikooli Eesti geenivaramu andmeid.

Coxi võrdeliste riskide mudelite korral, kus riskifaktoriteks on metaboliidid, polügeenseid riskiskoorid ja elustiililisi tunnused, on enne mudeli sobitamist kasutatud lassoregressiooni, mis tegi automaatse tunnuste valiku. Selgus, et selline lähene mine võimaldab küllaltki edukalt tuvastada olulisemad riskifaktorid ja vähendada mudeli keerukust.

Mudeli prognoosivõimeid hinnati Harrelli C-indeksi abil. Analüüsi tulemuste põhjal selgus, et metaboliitide, polügeensete riskiskooride ja elustiililise informatsiooni kaasamine prognoosivatesse mudelitesse tõepoolest suurendab mudeli headust iseloomustavat näitajat. Eriti hästi tuli see töös välja 2. tüüpi diabeedi puhul, mille korral nii meestel kui ka naistel C-indeksi 95% usaldusintervall ei kattunud teise kahe mudeli usaldusintervalliga ja seepärast saab väita, et antud mudeli sooritusvõime oli teisest kahest mudelist parem. Oluline järeldus on ka see, et naiste mudelid saavutasid üldisemalt kõrgema C-indeksi kui meeste omad.

Edasistes uuringutes võiks leitud mudelite usaldusväarsust ja üldistatavust testida sõltumatutes andmestikes. Ühtlasi saaks uurida vanuserühmade põhiseid mudeleid, et hinnata haigusriske erinevates vanusegruppides.

## Kasutatud kirjandus

- [1] Tervise Arengu Instituut. Rahvastiku tervise aastaraamat 2024. Eesti rahvastiku tervis ja selle mõjurid. Fookusteema: toitumine. *Tallinn: Tervise Arengu Instituut*, 2024.
- [2] OECD and Euroopa Komisjon. Health at a Glance: Europe 2016 - State of Health in the EU Cycle. *OECD Publishing, Pariis*, pages 4, 12, 177–190, 2016. doi:10.1787/9789264265592-en.
- [3] Lili Milani, Maris Alver, Sven Laur, et al. The Estonian Biobank's journey from biobanking to personalized medicine. *Nature Communications*, 16(3270), 2025. doi:10.1038/s41467-025-58465-3.
- [4] Eesti geenivaramu. Avati üle 200 000 geenidonorile mõeldud portaal MinuGeenivaramu. Vaadatud 20.04.2025. URL: <https://genomics.ut.ee/et/sisu/avati-ule-200-000-geenidonorile-moeldud-portaal-minugeenivaramu>.
- [5] Cardiovascular diseases, 2022. Vaadatud 24.04.2025. URL: <https://my.clevelandclinic.org/health/diseases/21493-cardiovascular-disease>.
- [6] World Health Organization. Cardiovascular diseases. Vaadatud 24.04.2025. URL: [https://www.who.int/health-topics/cardiovascular-diseases#tab=tab\\_1](https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1).
- [7] Tervisestatistika ja terviseuuringute andmebaas. SD15: Surmade esialgsed registreerimisandmed põhjuse, soo, maakonna ja kuu järgi, 2025. Vaadatud 24.04.2025. URL: [https://statistika.tai.ee/pxweb/et/Andmebaas/Andmebaas\\_\\_01Rahvastik\\_\\_04Surmad/SD15.px/](https://statistika.tai.ee/pxweb/et/Andmebaas/Andmebaas__01Rahvastik__04Surmad/SD15.px/).
- [8] Kristin Ly Järving, Tuuli Teeäär, Märt Elmet, et al. Kodade virvendusarütmia. 2021. Vaadatud 03.04.2025. URL: <https://www.kliinikum.ee/patsiendiinfo-andmebaas/kodade-virvendusarutmia/>.

- [9] Kodade virvendusarütmia. Vaadatud 08.04.2025. URL: <https://www.inimene.ee/haigused-ja-seisundid/list/haigused-ja-seisundid/kodade-virvendusarutmia>.
- [10] Dominik Linz, Monika Gawalko, Konstanze Betz, et al. Atrial fibrillation: epidemiology, screening and digital health. *The Lancet Regional Health - Europe*, 37:100786, 2024. doi:10.1016/j.lanepe.2023.100786.
- [11] Coronary artery diseases, 2024. Vaadatud 03.04.2025. URL: <https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/symptoms-causes/syc-20350613>.
- [12] British Heart Foundation. Global heart & circulatory diseases factsheet, 2025. Vaadatud 03.04.2025. URL: <https://www.bhf.org.uk/-/media/files/for-professionals/research/heart-statistics/bhf-cvd-statistics-global-factsheet.pdf>.
- [13] Tuuli Teeäär, Ülle Miitel, Elviira Vool, et al. Krooniline südamepuudulikkus, 2019. Vaadatud 04.04.2025. URL: <https://www.kliinikum.ee/patsiendiinfo-andmebaas/krooniline-sudamepuudulikkus/>.
- [14] Südamepuudulikkus. Vaadatud 04.04.2025. URL: <https://www.inimene.ee/haigused-ja-seisundid/list/haigused-ja-seisundid/sudamepuudulikkus>.
- [15] Heart failure, 2025. Vaadatud 04.04.2025. URL: <https://www.mayoclinic.org/diseases-conditions/heart-failure/symptoms-causes/syc-20373142>.
- [16] Tiina Uuetoa. Kroonilise südamepuudulikkuse epidemioloogia, käsitlus ja prognoos. *Eesti arst*, 98(Lisa 1):3–7, 2019.

- [17] Bianca Steiner, Anne Neumann, Yannick Pelz, et al. Challenges in heart failure care in four European countries: a comparative study. *European Journal of Public Health*, 33(3):448–454, 2023. doi:10.1093/eurpub/ckad059.
- [18] Diabeet. Vaadatud 24.04.2025. URL: <https://www.kliinikum.ee/diabeet/diabeet/>.
- [19] Eesti Diabeediliit. Mis on diabeet? Vaadatud 24.04.2025. URL: <https://www.diabetes.ee/mis-on-diabeet>.
- [20] World Health Organization. Diabetes. Vaadatud 24.04.2025. URL: [https://www.who.int/health-topics/diabetes#tab=tab\\_1](https://www.who.int/health-topics/diabetes#tab=tab_1).
- [21] Euroopa toiduohutusamet. Vaadatud 08.04.2025. URL: <https://www.efsa.europa.eu/et/glossary/metabolite>.
- [22] Ralf Tambets, Jaanika Kronberg, Erik Abner, et al. Genome-wide association study for circulating metabolites in 619,372 individuals. *medRxiv*, 2024. doi:10.1101/2024.10.15.24315557.
- [23] Jeffrey C. Barrett, Tõnu Esko, Krista Fischer, et al. Metabolomic and genomic prediction of common diseases in 700,217 participants in three national biobanks. *Nature Communications*, 2024. doi:10.1038/s41467-024-54357-0.
- [24] Malle Englas, Siiri Jakobson, Ebe Pilt, et al. Eesti vabariigi preemiad 2019. *Tallinn: Eesti Teaduste Akadeemia*, pages 95–97, 2019.
- [25] William S. Bush and Jason H. Moore. Chapter 11: Genome-wide association studies. *PLOS Computational Biology*, 8(12):1–11, 2012. doi:10.1371/journal.pcbi.1002822.
- [26] Wei Jiang, Ling Chen, Matthew J. Girgenti, and Hongyu Zhao. Tuning parameters for polygenic risk score methods using GWAS summary sta-

- tistics from training data. *Nature Communications*, 15(24), 2024. doi: 10.1038/s41467-023-44009-0.
- [27] Tian Ge, Chia-Yen Chen, Yang Ni, Yen-Chen Anne Feng, and Jordan W. Smoller. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, 10(1776), 2019. doi:10.1038/s41467-019-09718-5.
- [28] Kristi Läll. *Risk scores and their predictive ability for common complex diseases*. PhD thesis, Tartu Ülikool, 2019.
- [29] Great Learning Editorial Team. A Complete understanding of LASSO Regression, 2024. Vaadatud 25.05.2025. URL: <https://www.mygreatlearning.com/blog/understanding-of-lasso-regression/>.
- [30] David Collett. *Modelling Survival Data in Medical Research*. CRC Press, 2014.
- [31] Statistical Odds & Ends. What is Harrell’s C-index?, 2019. Vaadatud 21.05.2025. URL: <https://statisticaloddsandends.wordpress.com/2019/10/26/what-is-harrells-c-index/>.
- [32] SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *European Heart Journal*, 42(25):2439–2454, 2021. doi:10.1093/eurheartj/ehab309.
- [33] Rafael R. Oexner, Hyunchan Ahn, Konstantinos Theofilatos, et al. Serum metabolomics improves risk stratification for incident heart failure. *European Journal of Heart Failure*, 26(4):829–840, 2024. doi:10.1002/ejhf.3226.
- [34] Ruijie Xie, Sha Sha, Lei Peng, et al. Metabolomics data improve 10-year cardiovascular risk prediction with the SCORE2 algorithm for the general population without cardiovascular disease or diabetes. *medRxiv*, 2024. doi: 10.1101/2024.04.29.24306593.

## **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Gailin Savi,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose "Kardiovaskulaarsete haiguste ja 2. tüüpi diabeedi riskimudelite hindamine TÜ Eesti geenivaramu andmete põhjal", mille juhendajad on Mart Kals ja Krista Fischer, 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.

Gailin Savi

28.05.2025