

Tartu Ülikool
Peremeditsiini ja rahvatervishoiu instituut

SÜDAME-VERESOOKONNA HAIGUSTE
RISKIMUDELITE LOOMINE EESTI ELEKTROONILISTEL
TERVISEANDMETEL

Magistritöö rahvatervishoius

Laura Ilves

Juhendajad: **Taavi Tillmann, MD, PhD, Tartu Ülikool,**
peremeditsiini ja rahvatervishoiu instituut,
rahvatervishoiu kaasprofessor

Marek Oja, PhD, Tartu Ülikool, arvutiteaduse
instituut, terviseinformaatika teadur

Sirli Tamm, MSc, Tartu Ülikool, arvutiteaduse
instituut, programmeerija

Tartu 2024

Magistritöö tehti Tartu Ülikooli peremeditsiini ja rahvatervishoiu instituudis koostöös Tartu Ülikooli arvutiteaduse instituudiga. Uurimistöö on läbi viidud projektide „Kliiniliselt oluliste radade tuvastamine terviseandmetest“ (PRG1844) ning „Suurte terviseandmete ja sotsiaalandmete omavaheline integreerumine, tervise ebavõrdsuste paremaks mõistmiseks ja vähendamiseks“ (PSG809) raames.

Tartu Ülikooli rahvatervishoiu magistritööde kaitsmiskomisjon otsustas 24.05.2024 lubada väitekiri terviseteaduse magistrikraadi kaitsmisele.

Retsensent: Aet Saar, MD, PhD, Põhja-Eesti Regionaalhaigla, kardioloog; Tartu Ülikool, kliinilise meditsiini instituut, südamekliinik, kardioloogia teadur

Kaitsmine: 05.06.2024

Sisukord

Kasutatud lühendid	5
Lühikokkuvõte.....	7
1. Sissejuhatus	8
2. Kirjanduse ülevaade	9
2.1 Mõisted	9
2.2 Südame-veresoonkonna haiguste epidemioloogia.....	10
2.3 Südame-veresoonkonna haiguste riskitegurid	11
2.3.1 Vererõhk	11
2.3.2 Kolesterol.....	12
2.3.3 Rasvumine	13
2.3.4 Suitsetamine.....	13
2.3.5 Diabeet.....	14
2.4 Südame-veresoonkonna haiguste ennetamine	14
2.4.1 Mudelite täpsuse hindamine	15
2.4.2 Tarkvaralahendused riskimudelite loomiseks	16
2.4.3 Ülevaade senistest südame-veresoonkonna haiguste riskimudelitest.....	17
2.4.4 Riskimudelite piirangud	20
2.5 Südame-veresoonkonna haiguste uurimine elektrooniliste terviseandmete põhjal	20
2.5.1 Elektrooniliste terviseandmete kasutamise eelised SVHde uurimisel.....	20
2.5.2 Elektrooniliste terviseandmete kasutamise puudused	22
2.5.3 Teaduslik lünk	23
3. Eesmärgid	25
4. Materjal ja meetodika.....	26
4.1 Andmeallikad.....	26
4.2 Valimi moodustamine.....	27

4.3 Mudelite loomise metoodika ja andmeanalüüs.....	28
4.3.1 Mudelitesse kaasatud tunnused.....	28
4.3.2 Mudelite loomine ja andmeanalüüs	30
5. Tulemused.....	33
5.1 Tunnused uuringu algushetkel	33
5.2 Südame-veresoonkonna haigustesse esmashaigestumine ja suremus.....	34
5.3 Mudelite tulemused.....	36
5.3.1 Tegurite ennustusvõime	36
5.3.2 Diskrimineerimine	37
5.3.3 Reklassifitseerimise muutus.....	38
6. Arutelu	40
7. Järeldused ja ettepanekud	46
8. Kasutatud kirjandus	48
Summary	58
Tänuavaldus	60
Elulugu.....	61
Lisad.....	62
Lisa 1. Töös kasutatavad SVH definitsioonid diagnooside kaupa.....	62
Lisa 2. Mudeli tunnuste definitsioonid	64
Lisa 3. Mudelite koefitsiendid	65
Lisa 4. Riskikalkulatsioon kahe stsenaariumi näitel lõplikku mudelit kasutades	66
Lisa 5. Kohordi defineerimine tarkvaraga ATLAS	68

Kasutatud lühendid

AHA	Ameerika südameassotsiatsioon (<i>American Heart Association</i>)
ATC-kood	anatomilis-terapeutiline keemiline kood (<i>Anatomical Therapeutic Chemical code</i>)
CDM	ühtne andmemudel (<i>common data model</i>)
CI	usaldusvahemik (<i>confidence interval</i>)
DBP	diastoolne vererõhk (<i>diastolic blood pressure</i>)
eGFR	hinnanguline glomerulaarfiltratsiooni kiirus (<i>estimated glomerular filtration rate</i>)
ESC	Euroopa Kardioloogide Selts (<i>European Society of Cardiology</i>)
GBD	ülemaailmne haiguskoormuse uuring (<i>Global Burden of Disease study</i>)
HDL-C	kõrge tihedusega lipoproteiinide kolesterool (<i>HDL-cholesterol</i>)
HR	riskimäärade suhe (<i>hazard ratio</i>)
KMI	kehamassiindeks (<i>body mass index</i>)
KNH	krooniline neeruhaigus (<i>chronic kidney disease</i>)
LDL-C	madala tihedusega lipoproteiinide kolesterool (<i>LDL-cholesterol</i>)
MI	müokardiinfarkt (<i>myocardial infarction</i>)
MIR	Müokardiinfarkiregister (<i>Myocardial Infarction Register</i>)
NRI	reklassifitseerimise muutus (<i>net reclassification improvement</i>)
NHS	Ühendkuningriigi riiklik tervishoiuteenistus (<i>National Health Service</i>)
OHDSI	<i>Observational Health Data Sciences and Informatics</i>
OMOP	<i>Observational Medical Outcomes Partnership</i>
OR	šansisuhe (<i>odds ratio</i>)
PCE	<i>Pooled Cohort Equations</i>
RHK-10	rahvusvaheline haiguste klassifikatsioon, 10. versioon (<i>International Classification of Diseases</i>)

SBP	süstoolne vererõhk (<i>systolic blood pressure</i>)
SCORE	<i>Systematic Coronary Risk Evaluation</i>
SCORE-OP	<i>Systematic Coronary Risk Evaluation Older Persons</i>
SD	standardhälve (<i>standard deviation</i>)
SNOMED CT	meditsiiniterminid (<i>Systematic Nomenclature of Medicine Clinical Terms</i>)
SVH	südame-veresoonkonna haigused (<i>cardiovascular diseases</i>)
TIS	tervise infosüsteem (<i>health information system</i>)
UK	Ühendkuningriik (<i>United Kingdom</i>)
WHO	Maaailma Terviseorganisatsioon (<i>World Health Organization</i>)
ÄMI	äge müokardiinfarkt (<i>acute myocardial infarction</i>)

Lühikokkuvõte

Senine teadustöö südame-veresoonkonna haiguste (SVH) riski ennustamisel on keskendunud riskimudelitele, mis vajavad kliinilist visiiti ja vereproovi. On ebaselge, kui täpse riskihinnangu saab anda, kasutades ainult elektroonilisi terviseandmeid. Magistritöö põhieesmärk oli luua kaks riskimudelit, et ennustada esmase SVH tekkeriski tervete 40–74aastaste Eesti elanike seas, kasutades selleks elektroonilisi terviseandmeid. Töö alaeesmärkideks oli 1) kirjeldada SVH ja selle riskitegurite levimust kolme riikliku andmekogu põhjal loodud terviseandmebaasis; 2) uurida, milline on terviseandmetest tuvastatavate SVH riskitegurite eraldiseisev ennustusvõime; 3) mõõta loodud mudelite diskrimineerimist läbi sisemise valideerimise ja 4) võrrelda kahe loodud mudeli reklassifitseerimise muutust.

Riskimudelite loomiseks kasutati terviseandmete andmebaasi, mille aluseks on 10% juhuvalim Eesti rahvastikust ($n = 150\,824$), kelle kohta olid andmed olemas ajaperioodil 01.01.2012–31.12.2019. Uuringu valimi koostamiseks kaasati kõik SVH vabad 40–74aastased inimesed ($n = 52\,172$). Uuritavaid jälgiti esmase SVH tekke osas 5 aasta jooksul. Loodi kaks mudelit. Esimene mudel kaasas teguritena vanust ja sugu ning teine mudel kaasas lisaks vanusele ja soole veel üksteist tegurit. Mudelite loomiseks kasutati R-i paketti *PatientLevelPrediction*. Tulemuste kirjeldamiseks kasutati sagedustabeleid, seoste analüüsimiseks arvutati riskimäärade suhted (HR), pärast mudelite sisemist ristvalideerimist hinnati mudelite diskrimineerimist Harrelli C-statistiku meetodil ning arvutati reklassifitseerimise muutus (NRI).

5aastase jälgimisaja jooksul tekkis 1935 esmast SVH juhtu. Lõpliku mudeli C-statistik treeningandmetel oli 0,762 (95% usaldusvahemik CI 0,751–0,773) ja testandmetel 0,757 (95% CI 0,737–0,778). Diagnooside ja ravimite tarvitamise info kaasamine riskimudelisse parandas mudeli C-statistikut võrreldes algse vanus-soo mudeliga treeningandmetel +0,012 võrra ja testandmetel +0,010 ühiku võrra. Reklassifitseerimise muutus NRI oli 0,112 (95% CI 0,098–0,123), viidates lõpliku mudeli klassifitseerimise paranemisele võrreldes esialgse vanus-soo mudeliga. Kõige olulisemad ennustavad tegurid lõplikus mudelis olid meessugu (HR = 2,30; 95% CI 2,06–2,56), hüpertensiooni (HR = 1,54; 95% CI 1,37–1,73) ja kodade virvenduse ja laperduse (HR = 1,49; 95% CI 1,22–1,78) diagnoosid.

SVH riskitegurid olid Eesti elektroonilistest terviseandmetest tuvastatavad. Diagnooside ja ravimite tarvitamise info kaasamine riskimudelisse muutis mudeli ennustuse täpsemaks, mistõttu võiks tulevikus sõeluuringu sihtrühma defineerimisel lisaks vanusele ja soole arvesse võtta ka kaasuvaid haigusi ja ravimite tarvitamise infot.

1. Sissejuhatus

Südame-veresoonkonna haigused on rühm enneaegset suremust ja puuet põhjustavaid haiguseid, mis moodustavad ülemaailmselt kolmandiku surmadest (1). SVHga on seotud arvukalt riskitegureid, millest ühed olulisemad on kõrge vererõhk ja kolesterool, rasvumine, suitsetamine ning diabeet (2). Seoses rahvastiku vananemise ja kasvamisega on oodata riskitegurite laiemat levikut, mistõttu on vaja jätkuvaid jõupingutusi SVH ennetamiseks (1, 3). Hinnanguliselt 80–90% SVHdest on ennetatavad (4, 5), ainult muudetavate riskitegurite kontrollimisel on võimalik ennetada hinnanguliselt 52,6–57,2% SVHdest (2).

Ajalooliselt on suurem osa SVH ennetusest keskendunud meetmetele, et parandada kogu rahva tervisekäitumist. Paralleelselt üleriiklike meetmetega annab ka personaalsete teenuste arendamine kasu (6). Selleks, et paremini aru saada, kellel on kõrgem risk SVH tekkeks, on kasulikuks abivahendiks osutunud riskimudelid, mis võimaldavad erinevate riskitegurite põhjal ennustada, kellel on SVH teke kõige tõenäolisem (7–9). Riskimudeleid on võimalik luua erinevate andmete põhjal. Elektroonilised terviseandmed pakuvad selleks mitmeid eeliseid, sest võimalik on kaasata palju uuritavaid ning see on kulutõhusam, kuna ei eelda kliinilisi uuringuid ja protseduure (10–12).

Senine teadustöö SVH riski ennustamisel on keskendunud riskimudelitele, mis vajavad kliinilist visiiti ja vereproovi võtmist. Võrdlemisi vähe on uuritud, kui täpse riskihinnangu saab anda, kasutades ainult elektroonilisi terviseandmeid. Seni pole ükski teadusgrupp uurinud, kui täpseks saab teha esialgse vanus-soo mudeli eelhinnangu, lisades sinna usaldusväärset kogutud terviseandmed. Eestis pole uuritud, millised on SVH riskitegurite elektroonilistest terviseandmetest tuvastamise eripärad ja puuduvus ning kui täpne võib olla riskimudel, mis kasutab ainult elektroonilisi terviseandmeid.

Magistritöö annab ülevaate SVH riskitegurite levimusest ja eraldiseisvast ennustusvõimest kolme riikliku andmekogu põhjal loodud terviseandmebaasis, luues kaks riskimudelit esmase SVH ennustamiseks. Töö tulemused võimaldavad välja selgitada, kas loodud mudeli diskrimineerimine on aktsepteeritav ja kas lisategurite kaasamine riskimudelisse võib aidata paremini defineerida sõeluringuteenuse sihtrühma, mille alusel otsustada, keda kutsuda edasiste uuringutele.

2. Kirjanduse ülevaade

2.1 Mõisted

Diskrimineerimine (*discrimination*) on riskimudeli täpsuse hinnang, mis näitab, kui hästi mudel eristab tulemiga inimesi tulemita inimestest (13).

Elektroonilised terviseandmed (*electronic health records*) ehk digitaalne terviselugu on digitaalselt ja süstemaatiliselt kogutud info patsientide tervise kohta, mis ühendab erinevate tervishoiuasutuste infosüsteeme (14). Käesolevas töös kasutatakse elektroonilisi terviseandmeid üldmõistena, mille alla võivad kuuluda sõltuvalt riiklikest eripäradest info erinevatest registritest ja andmebaasidest.

Kalibreerimine (*calibration*) on riskimudeli andmestikule sobivuse hindamine ja sisuliste muudatuste tegemine mudeli ennustustes, et parandada prognooside täpsust (ehk vastavust tulemile) (13).

Linkimine (*data linkage*) on erinevatest allikatest pärineva teabe kõrvutamise protsess, et luua väärtuslik andmekogu (15). See võimaldab erinevatest registritest pärinevat infot ühe konkreetse isiku kohta omavahel siduda, et tekiks konstrueeritud ja kronoloogiline sündmuste jada. Käesolevas töös on linkimise all mõeldud protsessi, kus erinevatest allikatest pärit elektroonilised terviseandmed ühe konkreetse isiku kohta on omavahel seotud üheks terviklikuks andmebaasiks.

Riskimudel (*risk prediction model*) on statistiline mudel, mis kombineerib omavahel erinevate tegurite infot. Mudelitel võib olla erinevaid väljundeid, kuid kliinilise riskimudeli peamine eesmärk on täpselt tuvastada erinevatesse riskikategooriatesse kuuluvad inimesed. (13)

Südame-veresoonehaigused (*cardiovascular diseases*) on rühm haiguseid, mis kahjustavad südant ning veresooni (4), millest kõige sagedasemad on isheemiline südamehaigus, insult, perifeersete arterite haigus ning kodade virvendusarütmia (3). Laiem haiguste rühm on lähtuvalt rahvusvahelise haiguste klassifikatsiooni 10. versioonist (RHK-10) defineeritud diagnoosikoodidega I00–I99 (16). Isheemilise südamehaiguse ja insuldi puhul on olemas tõhusad ennetussekumised, mistõttu keskendutakse selles magistritöös peamiselt nendele haigustele.

Testandmed (*test set*) moodustavad valimi, mida kasutatakse treeningandmetel loodud riskimudeli testimiseks. (13)

Treeningandmed (*training set*) moodustavad valimi, mida kasutatakse riskimudeli loomiseks. (13)

2.2 Südame-veresoonkonna haiguste epidemioloogia

Südame-veresoonkonna haigused on jätkuvalt oluliseks rahvatervishoiu probleemiks, moodustades ülemaailmselt kolmandiku kõikidest surmadest (1). Maailma Terviseorganisatsioon (WHO) defineerib enneaegset suremust kui surma enne 70. eluaastat (17). 2019. aastal põhjustas SVH ülemaailmselt 38% enneaegsetest surmadest (18). Ülemaailmse haiguskoormuse uuringu (GBD) alusel suri SVH tõttu 2017. aastal 17,8 miljonit inimest, seejuures isheemiline südamehaigus ja insult moodustasid nendest surmadest kokku 85% (1).

Ajavahemikul 1990–2019 kasvasid ülemaailmsed SVH surmad 12,1 miljonilt 18,6 miljonini (1). Surmade kasvu seostatakse peamiselt 55% ulatuses rahvastiku vananemise ja 25% ulatuses rahvaarvu suurenemisega (3). Sama ajavahemiku jooksul on ülemaailmne SVH haigestumine peaaegu kahekordistunud – 271 miljonilt 523 miljonini (1). See viitab sellele, et surmade kasv võib ka tulevikus jätkuda.

Kuigi Euroopas on Euroopa Kardioloogide Seltsi (ESC) statistika järgi vanusele standarditud SVH suremus langenud, on see endiselt kõige levinum surmapõhjus. Igal aastal põhjustab SVH Euroopa regioonis, mille hulka arvatakse ESC alusel 56 riiki, umbes 3,9 miljonit surma, moodustades 45% kõikidest surmadest. Alla 70aastaste naiste seas põhjustas SVH 29% ja meeste seas 33% surmadest, millega on SVH Euroopas meeste seas juhtival kohal enneaegsete surmade põhjuste seas. Naiste seas on peamine enneaegse suremuse põhjus vähk, kuid SVH üldsuremuse näitajad on kõrgemad. Ajaperioodil 1990–2015 on SVH haigestumine ESC liikmesriikide seas kasvanud – meeste seas 4,3 miljonilt 5,3 miljoni juhuni ning naiste seas 4,9 miljonilt 5,7 miljoni juhuni. 2019. aastal tekkis ESC Euroopa piirkonna riikides ligikaudu 12,7 miljonit uut SVH juhtu, millest isheemiline südamehaigus moodustas naiste seas 45% ja meeste seas 39%. (3)

Nii nagu ülejäänud Euroopas, on ka Eestis SVH suremus langemas, kuid sellegipoolest on Eesti Euroopa üks kõrgema SVH suremusnäitajatega riike – 2020. aasta seisuga oli Eesti 8. kohal SVH suremuse poolest, jäädes endiselt kõrge riskiga riikide sekka (19). 2022. aastal suri Eestis vereringeelundite haigustesse (rahvusvahelise haiguste klassifikatsiooni RHK-10 koodid I00–I99) 8069 inimest, millest 39% moodustasid südame isheemiatõved (RHK-10 koodid I20–I25) ja peajuveresoonte haigused (RHK-10 koodid I60–I69) (20).

Eestis registreeritakse igal aastal ligikaudu 80 000 uut SVH juhtu (21). SVH põhjustab Eestis umbes pooled surmad (22). GBD uuringu alusel on Eestis SVH 2 korda sagedasem kui Euroopa keskmine ja 3 korda sagedasem kui maailma keskmine (23). Ka ESC riikide seas on Eesti kõrge sissetulekuga riikide seas SVH avaldumuse osas esikohal (1021 juhtu 100 000 inimese kohta) (3).

Kuigi SVH suremus on langenud, on haigestumine kasvanud, mis tähendab, et haiguse iseloom on muutunud. Kui varem oli SVH peamiselt surma põhjustav haigus, siis nüüd on see muutunud invaliidistavaks ja elukvaliteeti oluliselt langetavaks haiguseks (1, 23). Seoses muutustega rahvastiku tasandil on SVH ülemaailmseks ohuks ka tulevikus (3), mistõttu on haiguse ennetamisega vaja aktiivselt edasi tegeleda.

2.3 Südame-veresoonkonna haiguste riskitegurid

SVH riskitegurid võib üldistatult jagada kaheks – muudetavad ja mittemuudetavad. Mittemuudetavate tegurite hulka kuuluvad näiteks sugu, rass, vanus ja perekondliku SVH esinemise ajalugu (24). Muudetavaid riskitegureid on võimalik kontrollida õigeaegse ravi ja/või muudatustega tervisekäitumises. Sellisteks riskiteguriteks on näiteks ebatervislik toitumine, suurenenud kehakaal, suitsetamine, kõrgeenenud kolesterool ja vererõhk (1).

Eriti suur roll SVH tekkes on muudetavatel teguritel. Magnussen *et al* (2023) uuris viie muudetava riskiteguri (KMI, süstoolne vererõhk, LDL-kolesterool, suitsetamine, diabeet) mõju SVH 10aastasele avaldumusele ning leidis, et nende tegurite kontrollimisel oleks võimalik SVH teket ennetada vähemalt 57,2% naiste ja 52,6% meeste seas (2). Vanuse kasvades vähenevad seosed muudetavate riskiteguritega (25), kuid tõendus viitab, et vanemas eas SVH saamine ei ole vältimatu (26), mistõttu on ennetus oluline. Järgnevalt antakse ülevaade mõnest olulisemast muudetavast SVH riskitegurist.

2.3.1 Vererõhk

Ameerika Südameassotsiatsiooni (AHA) järgi peetakse kõrgvererõhktõveks püsivat vererõhu muutust, mida iseloomustab süstoolse vererõhu (SBP) väärtus ≥ 130 ja/või diastoolse vererõhu (DBP) väärtus ≥ 80 mmHg (27). ESC viimase definitsiooni järgi on kõrgvererõhktõve kriteeriumiteks SBP väärtus ≥ 140 ja/või DBP väärtus ≥ 90 mmHg (28). Seda erinevust selgitab asjaolu, et kõrgvererõhktõve piirmäärasid on aastakümnete jooksul toodud järjest madalamale, sageli viisil, kus Ameerikas muudetakse definitsioone varem kui Euroopas.

Ajavahemikul 1990–2019 suurenes seoses rahvastiku vananemise ja populatsiooni kasvuga inimeste hulk, kellel on kõrge vererõhk 2,18 miljardilt 4,06 miljardini (1). Eesti Müokardiinfarktiregistri (MIR) 2021. aasta aruande põhjal oli ägeda müokardiinfarkti (ÄMI) saanud patsientide seas 81,2%l kõrgvererõhutõbi (29).

Kõige madalam SVH risk on inimestel, kelle SBP väärtus on vahemikus 90–114 mmHg ning DBP väärtus vahemikus 60–74 mmHg (30). 40–70aastaste seas kahekordistab iga 20 mmHg SBP ja 10 mmHg DPB väärtuse tõus alates vererõhu väärtusest 115/75 mmHg SVH riski (31). Hüpertensiooniga inimestel on 30aastaselt risk elu jooksul saada SVH 63,3%, samas vanuses normaalsete vererõhu väärtustega inimeste seas on see risk 46,1% (30). 55–64aastaste seas on SBP väärtuse tõus 10 mmHg seotud 45% kõrgema riskiga saada isheemiline südamehaigus ja ligi 65% kõrgema riskiga saada isheemiline või hemorraagiline insult (31, 32). SVH kõrge riskile võib viidata vererõhu pikaajaline varieeruvus (33) ja see võib sõltuda ka SVH tüübist (30).

2.3.2 Kolesterool

Kolesterool on lipiid ehk rasv, mida leidub inimese vereringes. Normaalne üldkolesterooli näit on 5 mmol/l või madalam (34, 35). Kolesterool jaguneb madala tihedusega lipoproteiinide kolesterooliks (LDL-C) ja kõrge tihedusega lipoproteiinide kolesterooliks (HDL-C). LDL-C kleepub veresoonte seinte külge ning põhjustab ateroskleroosi (34, 35). Soovitatavat LDL-C taset on hiljuti langetatud. See peaks olema sõltuvalt inimese SVH riskitasemest vahemikus 1,4–2,6 mmol/l (8).

WHO andmetel esines 2008. aastal üle 25aastaste seas 39%l üldkolesterooli kõrgenemist ning umbes kolmandik isheemilise südamehaiguse või insuldi surmadest olid seostatavad kõrge LDL-C tasemega (36). Võrreldes teiste maailmajagudega oli 2019. aastal üldkolesterooli taseme $\geq 4,9$ mmol/l levimus kõrgeim Euroopas (54% mõlema soo puhul) (1, 36, 37). Eesti MIRi 2021. aasta aruandes esines ÄMI patsientide seas düslipideemiat 64,8%l (29). Kolesterooli ebanormaalne tase on seotud kõrgema SVH riskiga, eriti kõrge on isheemilise südamehaiguse tekkerisk. Näiteks meeste seas, kellel oli üldkolesterooli tase $\geq 6,2$ mmol/l, esines 2,2–3,6 korda suurem isheemilise südamehaiguse tekkerisk võrreldes meestega, kelle kolesterooli tase oli normaalsete väärtuste piires (38). Kõrgema üldkolesterooli tasemega meeste seas oli ka SVH suremus 2,1–2,9 korda suurem (38). SVH riski tõstavad nii normaalsetest väärtustest kõrgem üldkolesterooli tase (riskimäärade suhe HR = 1,2; 95% CI 1,2–1,3), kui ka LDL-C tase (HR = 1,2; 95% CI 1,1–1,4) (39). LDL-C vahemikus 4,1–4,9 mmol/l (HR = 1,7; 95% CI 1,4–2,2) ja $\geq 4,9$ mmol/l (HR = 1,5; 95% CI 1,2–2,1) seostuvad suurema SVH suremusega (40).

HDL-C ja SVH suremusriski vahel on leitud positiivseid seoseid – pikaajaline madalam LDL-C ja kõrgem HDL-C tase on seotud madalama SVH tekkeriskiga (8, 39, 41). 10% üldkolesterooli langust langetab isheemilise südamehaiguse tekkeriski 40aastaselt 50%, 50aastaselt 40% ja 60aastaselt 30% (42).

2.3.3 Rasvumine

WHO andmete järgi on ülemaailmne rasvumine kolmekordistunud alates 1975. aastast. 2016. aasta seisuga oli ülemaailmselt 39% täiskasvanutest ülekaalulised ning 13% rasvunud. 2019. aastal oli mittenakkushaiguste seas hinnanguliselt 5 miljonit rasvumisega seostatavat surma, mis moodustasid 12% mittenakkushaigustega seotud surmadest. (43)

Rasvumine on krooniline haigus, mida saab kategoriseerida kehamassiindeksi (KMI) järgi. $KMI \geq 25$ viitab ülekaalulisusele, $KMI \geq 30$ viitab rasvumisele (44). Rasvumine on SVH ja sellega seotud suremuse riskitegur. Paljud SVH riskitegurid, nagu näiteks kõrgeenenud üldkolesterool ja vererõhk, on levinud rasvunud inimeste seas (45).

KMI väärtuse ja suremuse vahel on leitud lineaarseid, J- ja U-kujulisi seoseid (46), sõltuvalt uuritava suitsetamise staatusest, ravimitest ja kaasuvatest haigustest (47). Lisaks on KMI ja vööümbermõõt tugevalt seotud SVH ja 2. tüüpi diabeediga (48). Riaz *et al* (2018) leidsid, et rasvumine on statistiliselt oluliselt seotud 2. tüüpi diabeedi (šansisuhe $OR = 1,67$; 95% CI 1,30–2,14) ja isheemilise südamehaigusega ($OR = 1,20$; 95% CI 1,02–1,41), kuid seos rasvumise ja insuldi vahel ei olnud statistiliselt oluline ($OR = 1,02$; 95% CI 0,95–1,09) (49). Rasvunute seas on hinnanguliselt 51%l kõrgvererõhktõbi, 8%l isheemiline südamehaigus, 3,5%l südamepuudulikkus ning varasem MI on olnud 21%l (50).

2.3.4 Suitsetamine

WHO andmete järgi sureb igal aastal tubakatarvitamise tõttu üle 8 miljoni inimese (51). Tubakas võib põhjustada surma ka mittesuitsetajate seas – passiivse suitsetamisega seostatavaid surmajuhtumeid on igal aastal ligikaudu 1,2 miljonit (51). Hinnanguliselt 10% SVHst on põhjustatud suitsetamise tõttu (52). Eesti MIRi 2021. aasta aruandes esines ÄMI patsientide seas suitsetamist 27,5%l, endiseid suitsetajaid oli 15,1% (29).

Suitsetamine põhjustab ligi 50% kõigist välditavatest surmajuhtumitest, kusjuures pooled neist on tingitud SVHst (51). Eluaegsel suitsetajal on 50%line tõenäosus surra suitsetamise tõttu ja keskmiselt kaotab ta 10 aastat oma elust (53). SVH risk alla 50aastastel suitsetajatel on viie- kuni

kümnekordne võrreldes mittesuitsetajatega (54, 55). Ka passiivne suitsetamine on seotud kõrgeenenud SVH riskiga (56).

Suitsetamine tõstab isheemilisse südamehaigusesse suremust võrreldes eluaegse mittesuitsetamisega 2–3 korda (57), risk on seotud ka suitsetamise pikkuse ja kogusega (58, 59). Isheemilise südamehaiguse patsientidel vähendab suitsetamisest loobumine suremust ja MI tekkeriski (60).

2.3.5 Diabeet

Diabeet on krooniline metaboolne haigus, mis põhjustab kõrgeenenud veresuhkru taset (61). Diabeedi tuntumad alatüübid on 1. ja 2. tüüpi diabeet, nendest viimast esineb diabeediga inimeste seas ligi 95%l (61). Võrreldes 1. tüübiga on 2. tüüpi diabeet ennetatav, näiteks toitumise muudatustega (62). WHO andmetel kasvas diabeedi esinemine 108 miljonilt 1980. aastal 422 miljonini 2014. aastal (61). Eesti MIRi 2021. aasta aruande järgi oli ÄMI patsientide seas diabeeti 24,4%l (29).

2. tüüpi diabeediga inimestel on ligikaudu 2–4 korda suurem risk saada SVH võrreldes mittediabeetikutega. Kõige kõrgemad riskid on seotud isheemilise südamehaiguse (HR = 2,0; 95% CI 1,83–2,19) ja isheemilise insuldiga (HR = 2,27; 95% CI 1,95–2,65). Risk oli kõrgem ka hemorraagilise insuldi (HR = 1,56; 95% CI 1,19–2,05); klassifitseerimata insuldi (HR = 1,84; 95% CI 1,59–2,13) ning ülejäänud vaskulaarsurmade osas (HR = 1,73; 95% CI 1,51–1,98). Riskimäärade suhted olid isheemilise südamehaiguse osas kõrgemad naiste, 40–59aastaste ning surmaga lõppevate SVHde osas. Hinnanguliselt 11% SVH surmadest on seostatavad diabeediga. (63)

2. tüüpi diabeet on raviga kontrollitav, kuid isegi, kui glükoositaseme veres on kontrolli all, tõstab see märkimisväärselt SVH riski. Esiteks võib diabeet põhjustada vaskulaarset kahju juba enne raviga alustamist või pole ravi piisav. Teiseks võib täheldada seda, et diabeetikutel on sageli riskitegurid, mis on ka SVH riskitegurid – näiteks vererõhk, suitsetamine, düslipideemia, rasvumine. (64, 65)

2.4 Südame-veresoonkonna haiguste ennetamine

Hinnanguliselt 80–90% SVHst on ennetatavad (4, 5). Ajalooliselt on suurem osa SVH ennetusest keskendunud mittepersonaalsetele meetmetele, et parandada kogu rahva tervisekäitumist (66). Paralleelselt üleriiklike meetmetega annab ka personaalsete teenuste arendamine kasu. Selliste teenuste osakaal muutub üha tähtsamaks, kui haiguse sagedus ajas langeb. Parim teadaolev mudel

antud teenuseks on sõeluuring. See on tõhus näiteks vähi personaalses ennetuses. SVH puhul pole tõhusat sõeluuringu teenust või selle analoogi maailmas välja arendatud. Selline teenus vajaks mitmeid omadusi, nagu näiteks piisavalt täpne test (või testide seeria), mis võimaldaks tuvastada, kes vajavad edasisi sekkumisi. Seoses sellega on tekkinud vajadus individuaalsemaks lähenemiseks (6). Rahvusvahelistes SVH ennetusjuhistes on jõutud ühisele arusaamale, et primaarse ennetuse üks oluline komponent on kõrge riskiga inimeste tuvastamine ja SVH ennetamine nii tervisekäitumise muutmise kui ravimitega (8, 67, 68).

Üks võimalus kõrge riskiga inimeste tuvastamiseks on riskimudelid, mis kaasates erinevaid riskitegureid võimaldavad tuvastada inimesed, kelle puhul on SVH teke tõenäolisem kui teistel. SVH ennetamise juhised soovitavad kasutada riskimudeleid, sest need aitavad tuvastada kõrge riskiga inimesed, mis võimaldab planeerida sellele järgnevat sobivaid sekkumisi nii patsiendi kui ühiskonna tasandil (7–9). Riskiennustus on mõeldud SVH tekke erapooletuks ning usaldusväärseks hindamiseks. Viiekümne aasta jooksul on loodud palju mudeleid, et ennustada SVH tekkeriski erinevates rahvastikes (6, 69). Riskimudelite eripäraks on see, et ennustavad tunnused võivad olla seotud tulemiga, kuid ei pea olema põhjuslikud (13). Enne mudeli rakendamist on oluline, et see oleks piisavalt täpne.

2.4.1 Mudelite täpsuse hindamine

Mudeli hindamiseks on mitmeid erinevaid võimalusi, mis põhinevad sellel, kui täpselt suudab mudel tulemit ennustada. Oluline on, et ennustusvõimet hinnatakse treeningandmetest eraldi testandmetel, kuna treeningandmetel mõõdetud tulemus võib olla kallutatud liigse optimismi suunas (ehk mudel on üle treenitud) (70). Mudeli täpsuse mõõtmiseks on kaks olulist dimensiooni: diskrimineerimine ja kalibreerimine.

Diskrimineerimine näitab, kui hästi mudel eristab neid, kellel tekib tulem, nendest, kellel ei teki tulemit (13). Mudel võib heterogeenses populatsioonis hästi eristada tulemiga inimesi tulemita inimestest, kuid sama mudel ei pruugi homogeensemas populatsioonis nii hästi töötada. Näiteks võib mudel, mis sisaldab ennustava tegurina vanust, töötada hästi 20–90aastaste populatsioonis, kuid ebaõnnestuda populatsioonis, kus on inimesed vanuses 50–60eluaastat.

Diskrimineerimist saab hinnata nii olukordades, kus kliinilise sekkumise piirmäär on defineeritud (nt ravimit soovitada/mitte soovitada) kui ka olukordades, kus see ei ole defineeritud (nt teavitada patsienti tema riskist). Kui piir on defineeritud, siis kasutatakse tundlikkust, spetsiifilisust, positiivset ja negatiivset prognoosväärtust. Ilma piiri määramata hinnatakse mudeli headust, kasutades kas ROC AUC meetodit või Harrelli C-statistikut. (13, 71)

C-statistik on nii tundlikkuse kui spetsiifilisuse mõõdik, mis väljendab tõenäosust, et juhuslikult valitud tulemiga inimesel on kõrgem absoluutne riskihinnang kui juhuslikult valitud tulemita inimesel. C-statistiku väärtused varieeruvad vahemikust 0,50 (juhuslikkus) kuni 1,0ni (täiuslik diskrimineerimine). Väärtuseid $< 0,70$ peetakse ebaadekvaatseteks, väärtuseid vahemikus $0,70-0,80$ aktsepteeritavateks ning väärtuseid $> 0,90$ suurepäraseks. (13)

Mudelit, mis ei tee hästi vahet kõrgema ja madalama riskiga inimeste vahel ehk mille diskrimineerimine on ebaadekvaatne, ei peaks kasutama. Diskrimineerimisest üksi ei piisa mudeli ennustusvõime hindamiseks ning hea diskrimineerimine ei taga head kalibreerimist. Mudeli teine dimensioon ehk **kalibreerimine** annab infot, kas mudeli ennustatud tõenäosus on vastavuses tegeliku sündmuse tõenäosusega. Kui hinnangud on täpsed, on mudel hästi kalibreeritud. Kalibreerimisest on kasu ainult siis, kui seda tehakse välisel valideerimisel. Sisemise valideerimise korral on kalibreerimine sageli näiliselt täiuslik. Ennustusmudel peaks pakkuma valideeritud tulemi ennustust uutele inimestele. Seega ei paku mudeli väljatöötamiseks kasutatud treeningandmed muud kui võimalust mudelile õppida. Seetõttu on valideerimine oluline protsess mudeli väljatöötamisel. (13, 71)

Sisemine valideerimine hindab mudeli valiidsust samal andmestikul, millel mudel loodi. Üks võimalus sisemise valideerimise läbi viimiseks on kasutada ristvalideerimist (*cross-validation*), mille käigus jagatakse andmestik juhuslikult n võrdseks osaks, kus n tuleb arvuliselt määrata (näiteks $n = 10$). Mudel treenitakse $n - 1$ alavalimil ning üks alavalim jäetakse valideerimiseks. Sellist meetodit korratakse iga alavalimiga, kokku valideeritakse mudelit n korda. **Väline valideerimine** on vajalik, et hinnata mudeli üldist rakendatavust. Kui sisemist valideerimist iseloomustab test- ja treeningandmete juhuslik jagamine, siis väline valideerimine sihhib inimesi, kes erinevad inimestest, keda kasutati mudeli treenimisel. Välise valideerimise korral peaksid inimesed pärinema teisest ajaperioodist, asukohast või uuringu erineva meetoodika poolest. (13, 71–73)

2.4.2 Tarkvaralahendused riskimudelite loomiseks

Terviseinfot võib leida väga erineval kujul, mis raskendab selliste andmete põhjal uuringute tegemist, eriti rahvusvahelisi võrdlusi ja mudelite valideerimist. Üks lahendus sellele probleemile on välja töötatud OHDSI (*Observational Health Data Sciences and Informatics*) kogukonna poolt, kes on loonud ühtse andmemudeli OMOP (*Observational Medical Outcomes Partnership*) CDM (*common data model*). Selle eesmärk on viia erineval algkujul olevad terviseandmed standardsele kujule. See lihtsustab oluliselt andmeanalüüsi ja võimaldab koostööd erinevate riikide vahel. (71)

OHDSI kogukond on loonud andmemudeliga sobituvaid erinevaid tarkvaralahendusi ja tööriistu. Näiteks võimaldab tööriist ATLAS OMOPi andmekujule viidud terviseandmete põhjal disainida epidemioloogilisi uuringuid ja analüüsida terviseandmeid. Selle abil saab defineerida huvipakkuvaid kohorte ja juhtumeid ning nende põhjal genereerida uuringutulemusi. ATLAS võimaldab teha ka patsiendipõhiseid ennustusi. Üheks selliseks töövahendiks on patsiendipõhine ennustuspakett *PatientLevelPrediction*, mis genereerib huvipakkuvate kohortide põhjal päringud andmebaasi, võimaldades läbi tarkvara R riskimudeli treenimist ja valideerimist. (71, 74)

2.4.3 Ülevaade senistest südame-veresoonkonna haiguste riskimudelitest

Kuigi SVH riskimudeleid on kirjanduse alusel vähemalt 363 (75), on vähesed neist kliinilises kasutuses. Neli enim kasutatud mudelit on SCORE2 (*Systematic Coronary Risk Evaluation*), SCORE2-OP (*Systematic Coronary Risk Evaluation Older Persons*), PCE (*Pooled Cohort Equations*) ja QRISK3. Kõik neli ennustavad nii fataalseid kui mittefataalseid juhtumeid. SCORE2, PCE ja QRISK3 ennustavad 10aastast SVH riski, SCORE2-OP võimaldab ennustada nii 5- kui 10aastast SVH riski.

SCORE2. ESC soovib alates 2021. aastast kasutada SVH ennustamiseks SCORE2 mudelit (8). Varem oli kasutusel SCORE, mis ennustas 10aastast surmaga lõppeva SVH riski. SCORE2 võtab arvesse ka surmaga mittelõppevaid juhtumeid. Mudel on loodud Euroopa rahvastikule ning sihtrühmaks on 40–69aastased terved, varasema SVH ja diabeedita inimesed. SCORE2 alusel on riigid jagatud nelja geograafilisse riskipiirkonda: väike, keskmine, kõrge ja väga kõrge risk. Eesti on kõrge riskiga riikide seas. (76)

SCORE2-OP. Mudel laiendab SCORE2 mudeli põhimõtteid vanemale populatsioonile. Mudel on suunatud kasutamiseks 70aastaste ja vanemate seas, et ennustada 5- ja 10aastast SVH riski neljas geograafilises riskipiirkonnas. Mudel kaasab samu tegureid, mis SCORE2, kuid mudel on kohandatud ja rek calibreeritud, et arvestada vananemisega seotud muutusi, mis võimaldab täpsemat riskiennustust vanemate inimeste seas. (77)

QRISK3. UKs on loodud QRISK3 riskimudel. See on loodud Inglismaa ja Walesi elektrooniliste terviseandmete põhjal ning selle eesmärk on ennustada 10aastast SVH haigestumisriski 25–84aastaste seas, välistades uuritavad, kellel on kas kaasnev SVH või kes juba tarvitavad statiinravimeid (78). Mudelit rakendatakse UK riikliku tervishoiuteenistuse (NHS) *Health Check* programmis (79), mis on sarnane SVH sõeluuringu teenusele, aga selle efektiivsust pole randomiseeritud kontroll-uuringuga hinnatud.

PCE. USAs on loodud PCE kalkulaator, mida soovitatakse AHA poolt ning sellega hinnatakse 10aastast SVH haigestumisrisi 40–79aastaste seas. (80)

Kõigis neljas riskimudelil kasutatakse ennustavate teguritena vanust, sugu, süstoolset vererõhku, suitsetamise staatust, üld- ning HDL-kolesterooli. SCORE2-OP, QRISK3 ja PCE kaasavad lisaks veel diabeedi staatust. QRISK3 kaasab kõige laiemat hulka tegureid, nagu näiteks KMI, kodade virvendus ja laperdus, krooniline neeruhaigus (KNH). Lisaks kaasatakse sellesse kalkulaatorisse elukoha andmeid. Tabelis 1 on välja toodud kõikide riskimudelite tunnused (vt Tabel 1).

Tabel 1. SCORE2, SCORE2-OP, QRISK3 ja PCE riskimudelite peamised tunnused (76–78, 80)

Tunnus	SCORE2	SCORE2-OP	QRISK3	PCE
Fataalse SVH definitsioon	I10–16, I20–25, I46–52, I60–69, I70–73, R96.0–96.1	I10–16, I20–25, I46–52, I60–69, I70–73, R96.0–96.1	–	–
Mittefataalse SVH definitsioon	I21–I23, I60–69	I21–I23, I60–69	I20–I25, I63–I64, G45	aterosklerootiline SVH*
Sihtrühm	40–69aastased	70aastased ja vanemad	25–84aastased	40–79aastased
Lisatingimused	puudub eelnev SVH ja 1. tüüpi diabeet	puudub eelnev SVH	puudub eelnev SVH, ei ole saanud statiinravi	puudub eelnev SVH
Kasutuspiirkond	Euroopa	Euroopa	UK	USA
Tegurid	vanus, sugu, süstoolne vererõhk, suitsetamise staatus, üldkolesterool, HDL-C	vanus, sugu, süstoolne vererõhk, suitsetamise staatus, üldkolesterool, HDL-C	vanus, sugu (meeste ja naiste mudel eraldi), süstoolne vererõhk, suitsetamise staatus, üldkolesterool, HDL-C	vanus, sugu, süstoolne vererõhk, suitsetamise staatus, üldkolesterool, HDL-C
Lisategurid	–	diabeet	KMI, rahvuslik/etniline staatus, elukoht, kodade virvendus ja laperdus, perekondliku SVH ajalugu, reumatoidartriit, diabeet	diabeet, rahvus, antihüpertensiivsete ravimite kasutamine
SVH juhtumite arv treening- andmestikus	30 121	10 089	meeste mudelis 203 016; naiste mudelis 160 549	2689
Statistilise mudeli tüüp	Fine ja Gray mudel	Fine ja Gray mudel	Coxi võrdeliste riskide mudel	Coxi võrdeliste riskide mudel

*Töö autorile teadaolevalt ei olnud tabeli loomise hetkel võimalik kirjandusest tuvastada, milliseid diagnoosikoode PCE mudel täpsemalt kaasas.

2.4.4 Riskimudelite piirangud

Igal riskimudelil on omad piirangud. Enamasti on mudelid loodud erinevatele rahvastikele. Kasutades selliseid mudeleid teistsugusel rahvastikel või riskirühmadel, võib see viia riski üle- või alahindamiseni (81). Seetõttu on oluline, et iga mudel oleks kohandatud vastavale rahvastikule, et oleks olemas kõige usaldusväärsem informatsioon, mille alusel sekkumisi planeerida (9, 81, 82).

Eestis on uuritud olemasolevate riskimudelite ennustusvõimet Eestis ja selle tulemusena selgus, et Eestis sobiks pärast kohandamist olemasolevatest SVH riskimudelitest kasutada PCE ja SCORE (SCORE2le eelnev versioon) mudeleid. Samas QRISK2 (QRISK3le eelnev versioon) Eesti tingimustes kasutamiseks ei sobiks, sest see on alakalibreeritud ehk alahindab haigestumise tõenäosust (83).

2.5 Südame-veresoonkonna haiguste uurimine elektrooniliste terviseandmete põhjal

Klassikalised epidemioloogilised uuringud annavad kõige täpsemat riskiteavet. Üheks probleemiks selliste uuringute puhul võib kujuneda madal osalejate määr, sest keeruline on uuringusse saada kõiki inimesi, mistõttu võib uuringu tulemuste üldistamine kogu rahvastikule keeruliseks osutada. Lisaks on probleemiks asjaolu, et osalejad on võrreldes üldrahvastikuga oluliselt tervemad. Näiteks on UK biopangas (*UK Biobank*) osalejate suremus 50% madalam võrreldes üldrahvastikuga (84). Kui kohort-uuringutel treenida mudeleid, võib see viia riski olulise alahindamiseni. Lisaks on kohort-uuringute läbi viimine sageli ajamahukas ning nendega kaasnevad suured kulud. Üks võimalus selliseid probleeme lahendada on uuringute läbi viimine, kasutades elektroonilisi terviseandmeid, kus on võimalik saavutada osalejate määr 95–100%. (84, 85)

2.5.1 Elektrooniliste terviseandmete kasutamise eelised SVHde uurimisel

Prognoos viitab meditsiini tähenduses erinevate haiguste või tervises seisunditega seotud sündmuste tekkeriskile. Primaarselt on fookuses olnud haiguskulu prognoosimine peale sümptomite teket või esmast diagnoosi (86). Samas on üha populaarsemaks muutumas haigusrisi hindamine tervete inimeste seas. Seega kõik prognoosid, mis arvestavad ka kaasuvaid haiguseid, muutuvad järjest olulisemaks, sest üha rohkem on inimesi, kes elavad ühe või rohkema haigusega (10). Prognoosid aitavad mõista erinevate haiguste koormust tervishoiusüsteemile ja rahvastikule. Hinnangud prognoosile on riskimudelitel seisukohalt väga olulised, sest aitavad kaardistada edasisi uurimissuundi ja sekkumisvõimalusi (86). Üha rohkem kasutatakse inimeste tervisega seotud

seisundite uurimiseks elektroonilisi terviseandmeid, mis pakuvad võrreldes klassikaliste uuringutega mitmeid eeliseid.

Korruga on võimalik uurida paljude erinevate tegurite seoseid tulemitega ilma osalejate määra languseta. Elektroonilised terviseandmed võimaldavad tuvastada palju suuremas mahus inimesi, kellel on huvipakkuv tervises seisund või haigus. See võimaldab teha üldistusi kogu rahvastiku kohta. On leitud, et elektroonilistest terviseandmetest tuvastatavad levimuse määrad on võrreldavad riiklikes uuringutes leitud määradega (11). Samas ei ole selliste uuringute puhul vaja uuringus osalejatega kontakteeruda, sest andmed on juba olemas, mistõttu on võimalik kaasata uuringusse kõik inimesed, kelle kohta leidub infot antud uurimisküsimuse raames.

Elektroonilised terviseandmed võimaldavad luua suuremahulisi ja esinduslikke kohorte. Näiteks on Inglismaa terviseinfosüsteemi põhjal loodud CALIBERi kohort ($n = 901\,629$). Selle kohordi põhjal on võimalik tuvastada andmeid vanuse, soo, vererõhu, suitsetamise staatuse, diabeedi ja statiinravimite kasutamise kohta (87). Lisaks on antud kohordi raames uuritud erinevate veremarkerite, ravimite kasutamise ja muude riskitegurite seoseid SVHga (88–90). Sellised kohordid on suuremahulised ja esinduslikud, mistõttu on nende andmete põhjal võimalik teha kvantitatiivseid üldistusi kogu rahvastiku kohta.

Kirjanduse alusel on leitud, et riskitegurite mõju SVH alatüüpidele võib olla erinev (91). Enamasti on omavahel lingitud info mitmetest andmebaasidest ja registritest, mistõttu on võimalik tuvastada seoseid mitmete tegurite ja tulemite vahel. Näiteks uuriti Inglismaal ligi 2 miljoni inimesega kohordis mitmete riskitegurite ja suitsetamise seoseid erinevate SVHdega ning leiti, et SVHd on võimalik elektroonilistest terviseandmetest edukalt tuvastada SVH alatüüpide spetsiifiliselt. Uuringu tulemusel selgus, et riskitegurite seosed võivad olla väga erinevalt seotud SVH alatüüpidega, mõjutades seeläbi riskiennustust (92).

Erinevate andmebaaside ja registrite omavaheline linkimine vähendab nihke tõenäosust.

Tavaliselt on elektrooniliste terviseandmete näol toimunud linkimine mitme andmebaasi vahel ja see annab tunduvalt täpsema uuringutulemuse võrreldes ainult ühe andmebaasi kasutamisega (12). Uuringud viitavad, et analüüsid andmeid ühe andmebaasi/registri põhisealt, võib see viia alahindamiseni. Näiteks võib ainult ühe andmebaasi kasutamine põhjustada ÄMI alahindamist 25–50% ulatuses võrreldes kolme omavahel lingitud andmebaasi kasutamisega (93). Seega andmebaaside linkimine on vajalik, et saada terviklikum ülevaade inimese teekonnast meditsiinisüsteemis.

2.5.2 Elektrooniliste terviseandmete kasutamise puudused

Üks olulisemaid probleeme, mis on seotud elektrooniliste terviseandmete kasutamisega, on puuduvad andmed (94–96). Sterne jt (2009) alusel eristatakse peamiselt kolme tüüpi puuduvaid andmeid (97):

1. Täiesti juhuslik puudumine (*missing completely at random*). Puuduvate ja vaadeldavate väärtuste vahel ei ole süstemaatilisi erinevusi. Näiteks võivad vererõhu mõõtmised puududa mõõtevahendi rikke tõttu.
2. Juhuslik puudumine (*missing at random*). Mis tahes süstemaatilist erinevust puuduvate väärtuste ja vaadeldavate väärtuste vahel saab seletada erinevustega vaadeldavates andmetes. Näiteks puuduolevad vererõhu väärtused võivad olla madalamad kui mõõdetud vererõhu väärtused, aga ainult seetõttu, et noortel inimestel on tõenäolisemalt vererõhu mõõtmised puudu. Kuna vanus on tugev vererõhu ennustaja, saab seda puuduvust korrigeerida, kui vanust kaasata vererõhu imputeerimisse.
3. Mittejuhuslik puudumine (*missing not at random*). Isegi pärast vaadeldavate andmete arvesse võtmist jäävad puuduvate väärtuste ja vaadeldavate väärtuste vahele süstemaatilised erinevused. Näiteks kõrge vererõhuga inimestel võib olla suurema tõenäosusega ära jäänud arsti külastusi, sest neil on peavalud. Kui ka peavalu staatus on uurijatele teadmata, ei saa nad seda tegurit kasutada vererõhu imputeerimiseks. Imputeeritavad vererõhu väärtused jäävad liiga madalaks nende seas, kellel tegelikult oli peavalu, aga peavalu jäi tunnuseks märkimata. (97)

Kõige olulisem on tuvastada andmete puudumise põhjused. Täiesti juhusliku puudumise korral ignoreeritakse üldjuhul andmete puudumist. See võib tähendada, et tegemist on esialgu kavandatust veidi väiksema valimiga. Samas on lubatud esindusliku valimi analüüsiks sobivad töövõtted. Juhusliku puudumise korral ei tohiks andmete puudumist ignoreerida, sest see võib endaga kaasa tuua vigased ja nihkega hinnangud uuritavatele suurustele. Kõige raskem olukord on siis, kui andmete puudumine on mittejuhuslik, sest selle lahendamiseks on väga piiratud võimalused. (97)

Imputeerimine (*imputation*) on kõige tavapärasem lähenemine puuduvate andmete probleemile, mille eesmärk on luua mitmeid erinevaid usutavaid ja kombineeritud andmekogumeid. Imputeerimist saab näiteks kasutada olukorras, kui uuritakse süstoolse vererõhu seost SVHga ja mõnedel inimestel puuduvad andmed süstoolse vererõhu kohta. Tõenäosus, et süstoolne vererõhk puudub, väheneb tõenäoliselt vanusega, sest arstid mõõdavad seda rohkem

vanematel inimestel. Kui eeldada, et andmed puuduvad juhuslikult ja on olemas vanuse andmed, siis saab üldise seose hindamiseks puuduolevad süstoolse vererõhu väärtused imputeerida. (97)

Seega võib mitmene imputeerimine parandada uuringute valiidsust. Samas ei tohiks imputeerimist rakendada automaatselt kõikide puuduvate tunnuste puhul. Iga tekkinud olukorra puhul peab puuduvate andmete struktuuri ja olemust analüüsima, et hinnata, kas imputeerimist on vaja ja võimalik edukalt teostada. (97)

Erinevate tegurite tuvastamist elektroonilistest terviseandmetest mõjutab see, kuidas terviseinfot dokumenteeritakse ja kogutakse. Info tuvastamisel võivad probleeme tekitada tegurid, mida ei koguta struktuurselt ning mille osas puudub ühtne dokumenteerimise viis. Üheks näiteks on suitsetamise staatus. Pujades-Rodriguezi *et al* uuringu alusel oli 27%l uuritavatest puudu suitsetamise staatus (92). See võib tuleneda sellest, et suitsetamise staatust pannakse kirja väga erineval moel, mistõttu on selliseid andmeid keeruline elektroonilistest terviseandmetest tuvastada. Teine võimalus on see, et selline info ei jõuagi erinevatel põhjustel andmebaasi. Näiteks võib see juhtuda, kuna puudub ühtne struktuurne vorm, kuhu selline info kirja panna.

Suitsetamise staatuse puhul on oluline ka see, kuidas sellist infot elektroonilistest terviseandmetest tuvastatakse. Kukhareva *et al* uuringus uuriti terviseandmete põhjal suitsetamise staatuse tuvastamist ja leiti, et läbilõike uuringutüübi puhul on üle 80% suitsetamise andmetest ebatäpsed (98). Seetõttu kasutati uuringus võrdluseks pikaajalisi andmeid. Võrreldes läbilõikeliste andmete kasutamisega võimaldas pikaajaliste andmete kasutamine tuvastada 49,4% rohkem suitsetamise staatust (98). Selle uuringu tulemused näitavad, et 24tunnine ajaaken suitsetamise staatuse tuvastamiseks võib olla liiga kitsas. Kui arst on lisanud suitsetamise kirje mõni aasta tagasi, siis on see info inimese suitsetamise staatuse kohta suure tõenäosusega relevantne ka tänasel päeval, mis selgitab ka seda, et pikaajaliste andmete kasutamine parandas oluliselt suitsetamise staatuse tuvastamist. Seega on elektrooniliste terviseandmete põhistes uuringutes oluline arvesse võtta andmete kogumise iseärasusi, sest need võivad oluliselt mõjutada uuringu tulemusi (99).

2.5.3 Teaduslik lünk

Senine kirjandus keskendub kahele SVH tekkeriski hindamise etapile. Esimene etapp on kasutada ainult vanust ja sugu, et eelhinnata SVH tekkeriski. Selle põhjal kutsutakse edasistele uuringutele ehk teise etappi vaid neid, kelle risk on kõrgem (näiteks üle 40aastased mehed ja üle 50aastased naised). Seda tehakse näiteks kliinilises vastuvõtus UK NHS *Health Check* programmi raames (79).

Esimest etappi saaks teha oluliselt täpsemaks, kui lisaks vanusele ja soole kaasata ka teisi terviseandmetest tuvastatavaid riskitegureid. Seni on vaid üks meeskond loonud SVH riskimudeleid, kasutades selleks ainult elektroonilisi terviseandmeid (QRISK). Sealsetes mudelites on kolesterool olnud enamikul uuritavatest puudu. Niivõrd suuremahuline imputeerimine võis tagantjärele olla ebatäpne ning see võis langetada riskimudeli vastuvõetavust ja usaldusväärsust (100). Lisaks on Inglismaal paigutatud QRISKi mudeli kasutamine teenuse lõppetappi, kus viimases riskihindamise etapis tehakse lõplik otsus sekkumise vajaduse kohta. Kuna kolesterool on suuresti terviseandmetest puudu, ei kasuta NHS *Health Check* programm elektroonilisi terviseandmeid üldse, et selekteerida, keda kutsuda kliinilistele vastuvõttudele, ja teeb seda vaid vanust ning sugu arvestades (79).

Seni pole ükski teadusgrupp eeldanud, et lõplik riskihinnang tehakse hiljem (kui kolesterool jmt on määratud) ja küsinud, kui täpseks saame viia esialgse vanus-soo mudeli eelhinnangu, kui sinna lisada vaid terviseandmed, mis on usaldusväärselt kogutud ja millel on madal puuduvus. Teoreetiliselt võiks selline mudel olla parema diskrimineerimisega kui ainult vanus-soo mudel, säilitades paralleelselt hea vastamismäära. See võiks aidata paremini sihtida SVH sõeluuringu teenuseid neile, kes kõige tõenäolisemalt vajavad kolesterooli- ja teisi mõõtmisi. Seni pole Eestis uuritud, millised on võimalused elektroonilistest terviseandmetest tuvastada SVH riskitegureid ja haigusstaatust, millised on andmete tuvastamise eripärad ja nende puuduvus. Samuti pole teada, millist eraldiseisvat ennustusvõimet omavad SVH riskitegurid ja kui täpne oleks riskimudel, mis kasutaks ainult elektroonilisi terviseandmeid. Selle välja selgitamine aitaks luua aluse riigipõhisele kulutõhusale SVH ennetussüsteemile Eestis.

3. Eesmärgid

Töö peamine eesmärk oli luua kaks südame-veresoonkonna haiguste riskimudelit, et ennustada esmase SVH tekkeriski tervete 40–74aastaste Eesti elanike seas, kasutades selleks elektroonilisi terviseandmeid ajaperioodist 01.01.2012–31.12.2019. Sellest lähtuvalt püstitati järgnevad alaeesmärgid:

- 1) kirjeldada SVH esmashaigestumist ja selle riskitegurite levimust kolme riikliku andmekogu põhjal loodud terviseandmebaasis;
- 2) uurida, milline on terviseandmetest tuvastatavate SVH riskitegurite eraldiseisev ennustusvõime;
- 3) luua kaks riskimudelit esmase SVH juhu ennustamiseks 40–74aastaste eelneva SVHta inimeste seas ja mõõta loodud mudelite diskrimineerimist läbi sisemise valideerimise;
- 4) võrrelda kahe loodud mudeli reklassifitseerimise muutust.

4. Materjal ja meetodika

Magistritöö valmis projekti „Valdkondliku teadus- ja arendustegevuse tugevdamine“ (RITA) tegevus 1 „Strateegilise TA tegevuse toetamine – „Masinõppe ja AI toega teenused““ raames, mille tegemiseks on Tartu Ülikooli inimuuringu eetikakomitee luba 300/T-23 ja Eesti Bioeetika ja inimuuringu nõukogu luba (1.1-12/653).

4.1 Andmeallikad

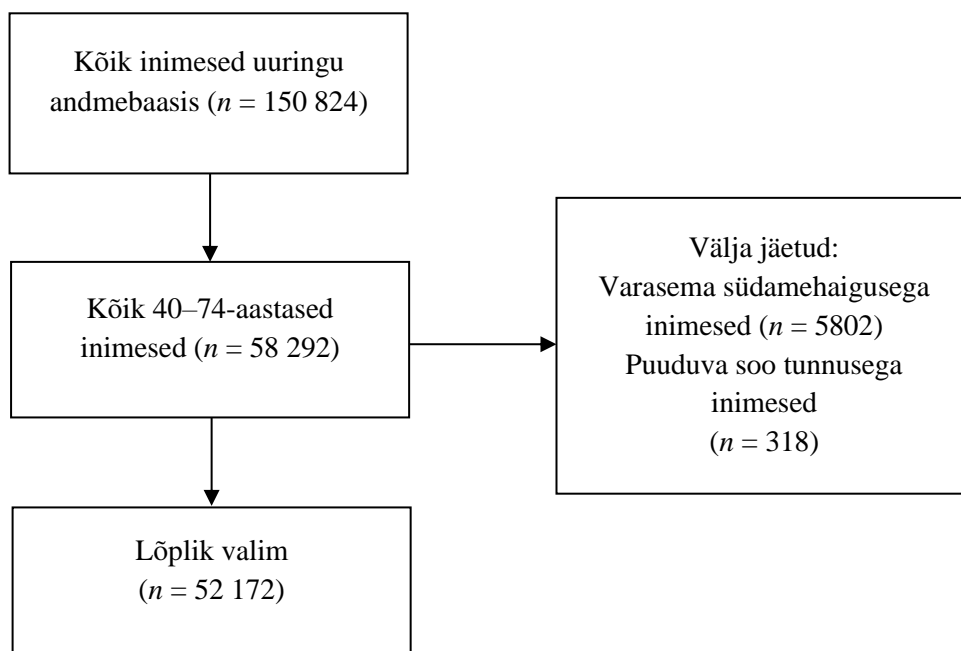
Elektrooniliste terviseandmete kasutamise võimalused Eestis on väga mitmekülgsed – kuigi nende peamine eesmärk on patsiendi terviseandmete infovahetus erinevate tervishoiuteenuste osutajate vahel, on selle teiseks väljundiks ka uuringute tegemine (101). Riskimudelite loomiseks kasutati varem loodud terviseandmete andmebaasi, mille aluseks on 10% juhuvalim Eesti rahvastikust ($n = 150\,824$), kelle kohta on andmed olemas ajaperioodil 01.01.2012–31.12.2019. Kõik isikuandmed on andmebaasis pseudonümiseeritud kujul ning neid ei ole võimalik seostada konkreetsete isikutega. Andmebaasi loomisel kasutati OMOP CDM andmemudelit ning andmed pärinesid kolmest riiklikust andmekogust – Retseptikeskusest, Tervisekassa raviarvete andmekogust ja Tervise infosüsteemist (TIS) (102).

TISist pärinevad andmed hõlmavad nii statsionaarseid kui ambulatoorseid kokkuvõtteid patsientide haiguslugudest, sisaldades nii haigusloo ajalugu, diagnoose, laborianalüüse ja nende tulemusi, protseduure, ravimeid ja allergiaid, suunamisi ja nende vastuseid. **Retseptikeskusest** pärinevad andmed sisaldavad väljakirjutatud ja -ostetud ravimite nimetusi ja koode, anatoomilisi-terapeutilisi koode (ATC-koodid), koguseid, tarvitamise juhiseid, väljaostu kuupäeva ja asukohta, ravimi määranud tervishoiuteenuse osutaja andmeid ning retseptiga seotud diagnoose. **Tervisekassa raviarvete andmekogust** pärinevad andmed sisaldavad diagnoose, osutatud teenuseid (sh vastuvõttud, laborianalüüsid, protseduurid ja ravimite manustamise info) ning kirurgilisi protseduure.

Ükski nimetatud andmekogudest ei ole esmane suremusandmete allikas, surmasid on andmebaasis olemas ligikaudu 67% (102). Teisisõnu, antud andmebaasis on 33% surmadest puudu, mis langetab fataalsete juhtude usaldusväärust ja mille tõttu antud töösse fataalseid juhtumeid ei kaasatud. Demograafilistest andmetest on andmebaasis olemas inimese sugu ja sünniaasta, kuid andmebaasist puuduvad Rahvastikuregistri andmed, mistõttu polnud võimalik välja selgitada inimesi, kes ei olnud kohordis jälgimisaja lõpuni.

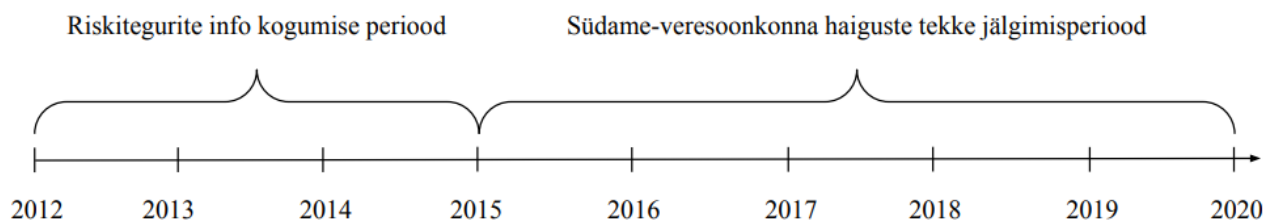
4.2 Valimi moodustamine

Andmebaasi kasutades viidi läbi retrospektiivne kohort-uuring. Magistritöösse kaasati kõik mehed ja naised, kes olid seisuga 1. jaanuar 2015 40–74aastased. Sellele kriteeriumile vastas 58 292 inimest. Kohordist eemaldati isikud, kellel oli seisuga 31. detsember 2014 diagnoositud eelnev südame-veresoonkonna haigus või kellel polnud andmetest tuvastatav sugu. Eelnevaks SVHks loeti vähemalt ühe RHK-10 koodi esinemine järgnevast loetelust: I20–25, I60–66, I69, I67,2, I70, I73,9, G45. Pärast eelneva SVHga ($n = 5802$) ja puuduva soo tunnusega inimeste ($n = 318$) eemaldamist jäi kohorti 52 172 inimest. Valimi koostamise voodiagramm on esitatud Joonisel 1.



Joonis 1. Valimi koostamise voodiagramm.

Kohordis olevaid inimesi jälgiti kuni esmase SVH tekke või uuringu lõpuni (31. detsember 2019). Tulemiks oli esmane isheemiline SVH, mis defineeriti läbi südame isheemiatõve, isheemilise insuldi või infarkti. Inimesel oli esmane SVH, kui selle kohta esines vähemalt üks RHK-10 kood järgnevast loetelust: I21–25, I63–64. Esmasjuhuks loeti jälgimisaja jooksul kõige varasem dokumenteeritud SVH juhtum. Nii eelneva kui ennustatava SVH detailsemad definitsioonid on välja toodud Lisas 1. Riskitegurite infot koguti 3 aasta jooksul (01.01.2012–31.12.2014) enne jälgimisperioodi algust. Uuringu ajajoon on kujutatud Joonisel 2.



Joonis 2. Uuringu ajajoon.

4.3 Mudelite loomise meetodika ja andmeanalüüs

4.3.1 Mudelitesse kaasatud tunnused

Magistritöö raames loodi 2 mudelit. Esimene mudel sisaldas teguritena vanust ja sugu. Teine ehk lõplik mudel sisaldas lisaks vanusele ja soole veel üksteist tegurit (Tabel 2). Mudelitesse kaasatud tunnuste detailsemad definitsioonid on välja toodud Lisas 2.

Demograafiliste tunnuste hulgast kaasati inimeste vanus ja sugu. Sugu käsitleti binaarse tunnusena. Naised võeti mudelis võrdlusrühmaks (ehk riskimäärade suhe $HR = 1$). Vanust käsitleti mudelis pideva tunnusena. Vanusevahemiku 40–74 valikul lähtuti sellest, et mitmetes riikides on SVH ennetamise sihtrühmaks üle 40aastased mehed ja üle 50aastased naised. Kuna mudel loodi mõlema soo jaoks kokku, lähtuti alumise vanusepiiri valikul madalamast väärtusest. Ülemise vanusepiiri valiku osas lähtuti teistes mudelites kasutatud ülemisest lävest 74 eluaastat. Samuti rakendatakse sellist vanusepiiri UKs läbi viidavas NHS *Health Check* programmis (79).

Diagnoosid ja ravimite tarvitamine. Mudelisse kaasatavaid diagnoose ja ravimite tarvitamist käsitleti binaarsete tunnustena, kus diagnoosi ja ravimite tarvitamise võrdlusrühmaks oli diagnoosi puudumine või ravimi mittetarvitamine. Arvesse läksid enne jälgimisperioodi algust diagnoositud haigused. Ravimite tarvitamine läks arvesse, kui vastava toimeainega ravimit oli enne jälgimisperioodi algust välja ostetud vähemalt kahel korral.

Diagnoosid ja ravimite tarvitamine valiti välja kirjanduse alusel ja eelnevates riskimudelites kasutatud tunnuste järgi. Tingimused olid järgnevad:

1. Mudelisse valitavad tegurid on eelmistes mudelites olnud SVHd ennustava efektiga ehk nende riskimäärade suhe on olnud $HR > 1$ või $HR < 1$ ning p -väärtus vähemalt ühes mudelisis $< 0,05$.
2. Magistritöö raames oli mudelite loomiseks kasutatav valim väiksem võrreldes varasemate mudelite aluseks olevate valimitega, seetõttu pidi kaasatavate haiguste levimus eelnevates mudelites olema vähemalt 0,3% nii naiste kui meeste mudelites.

3. Ravimite tarvitamisest kaasati eelnevate mudelite põhjal kortikosteroidide tarvitamine. Lisaks kaasati antidepressantide ning uinutite ja rahustite tarvitamine, et hinnata nende mõju SVH riskimudelil, kuna kirjanduse alusel on leitud erinevaid seoseid SVHga (103,104).
4. Kuna kroonilise neeruhaiguse raskusastme diagnoose (RHK-10 alusel koodid N18.3–18.5) kasutades ei olnud võimalik kaasata ühtegi uuritavat, lisati sellele diagnoosile lisakriteerium, mille alusel kaasati kõik inimesed, kellel on kirjas vähemalt kaks hinnangulise glomerulaarfiltratsiooni kiiruse (eGFR) mõõtmistulemust < 60 mL/min/1,73m² vähemalt kolme kuu pikkuse vahega (105,106).

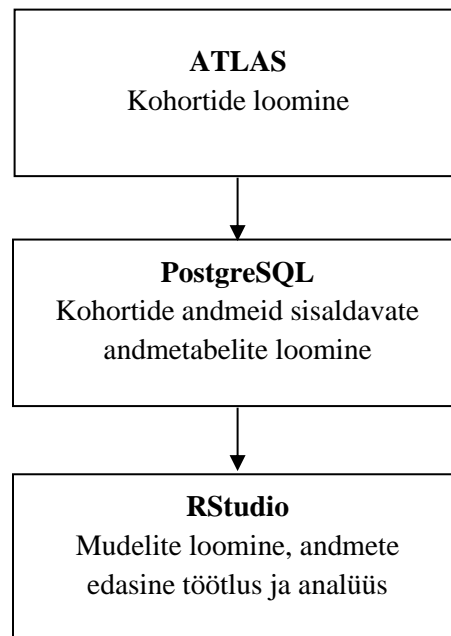
Tabel 2. Mudelitesse kaasatud tunnuste loetelu. Vastavasse mudelisse kaasatud tunnused on tähistatud märkega „x“

Tunnus	Mudel 1 Vanus-soo mudel	Mudel 2 Lõplik mudel
Demograafilised tunnused		
Sugu	x	x
Vanus	x	x
Diagnoosid		
Hüpertensioon		x
Kodade virvendus ja laperdus		x
Krooniline neeruhaigus		x
Migreen		x
Raske vaimne haigus		x
Reumatoidartriit		x
1. tüüpi diabeet		x
2. tüüpi diabeet		x
Ravimite tarvitamine		
Antidepressandid		x
Kortikosteroidid		x
Uinutid ja rahustid		x

Lõplikusse mudelisse kaaluti lisaks vanusele ka teiste pidevate tunnuste kaasamist. Pidevate tunnuste kaasamisele seati kriteerium, mille alusel ei kaasatud tegureid, mida ajavahemikul 01.01.2012–31.12.2014 ehk riskitegurite kogumise perioodil esines vähem kui 50%l isikutest. Sellele kriteeriumitele ei vastanud ükski pidev tunnus järgnevatest, mille esinemissagedus kohordis on välja toodud sulgudes: kehamassiindeksi väärtus (8,0%); üldkolesterooli mõõtmistulemus (26,0%) ja süstoolse vererõhu mõõtmistulemus (26,2%).

4.3.2 Mudelite loomine ja andmeanalüüs

OMOP tüüpi andmestikud sisaldavad standardiseeritud rahvusvahelisi meditsiinitermineid (SNOMED CT) (107). Magistritöös kasutati kõikidele RHK-10 koodidele vastavaid SNOMED CT termineid. Nende põhjal moodustati ATLASe tööriista kasutades mudelisse kaasatavate tunnuste kohordid (Lisa 5), mis sisaldavad kaasatavate tunnuste definitsioone. Kohortide andmeid sisaldavate andmetabelite loomiseks kasutati PostgreSQL-i. Mudeli loomine, andmete analüüs ja edasine töötlus toimus tarkvaraga R. Andmetöötluste protsessi illustreerib Joonis 3.



Joonis 3. Magistritöö andmetöötluste protsess.

Mõlema mudeli loomiseks kasutati Ri paketti *PatientLevelPrediction*, versioon 6.3.7 (74). Mudeli tüübina kasutati Lasso Coxi regressiooni. Lasso on lisafunktsioon, mis võimaldab mudeli piires tuvastada vähem olulised tegurid, jättes need mudelist välja (71). Mudelid jaotati vastavalt paketi algseadetele 75% treening- ja 25% testandmeteks. Mudelitele teostati 10kordne sisemine ristvalideerimine.

Coxi regressiooni abil arvutati riskimäärade suhted iga mudelisse kaasatava teguri kohta. Mudelid andsid lõpptulemusena iga teguri kohta koefitsiendi ehk hinnangu. Koefitsientide põhjal arvutati välja riskimäärade suhted, mille arvutamisel lähtuti järgnevast valemist $HR = e^{\text{koefitsient}}$.

SVHd kirjeldavate tegurite esitamiseks kasutati sagedustabeleid koos absoluutarvude, suhteliste sageduste ja p-väärtustega. Sagedustabelites kasutati rühmade võrdlemiseks vastavalt juhtude arvule kas hii-ruut testi või Fisheri testi. Olulisuse nivooks seati 0,05.

Magistritöö käigus võrreldi omavahel kahte mudelit – vanus-soo mudelit (mudel 1) lõpliku mudeliga (mudel 2). Võrdluse jaoks arvutati välja kategooriline reklassifitseerimise muutus (NRI). Kategooriline NRI on mõõdik, mis arvestab patsientide liikumist üles ehk kõrge riskihinnangu kategooriasse või alla ehk madala riskihinnangu kategooriasse. NRI arvutamisel testitakse hüpoteesi, kus

$$H_0: NRI = 0$$

$$H_1: NRI \neq 0,$$

seega eesmärk on hinnata, millisel määral erineb mudel 2 riskihinnang mudel 1 hinnangust. Kui NRI on positiivne, viitab see mudel 2 paremusele. Kui NRI on negatiivne, siis viitab see mudel 1 paremusele.

Lähtuvalt SVH ennetamise juhistest, klassifitseeriti patsiendid kõrge ja madala riskiga patsientideks. Kõrge risk on vastavalt juhiste risk > 10% saada 10 aasta jooksul SVH (8). Kuna andmestik võimaldas jälgimisaja pikkust kuni 5 aastat, kohandati riskimäär sellega sobivaks. Uuritavad määrati kõrge riskiga rühma, kui nende 5aastane risk saada esmane SVH oli 5% või kõrgem. Selline piirmäär võimaldab vastavalt Euroopa ja Eesti ravijuhistele ligikaudselt hinnata, kas inimesele soovitada või mitte soovitada kolesterooli alandavate ravimite tarvitamist. NRI välja arvutamiseks kasutati järgnevat valemit:

$$NRI = O(\text{üles}/\text{tulem}) - O(\text{alla}/\text{tulem}) + O(\text{alla}/\text{tulemita}) - O(\text{üles}/\text{tulemita}),$$

kus *üles* kirjeldab olukorda, kus inimese individuaalne riskihinnang tõuseb uut mudelit (mudel 2) kasutades, võrreldes vana mudeliga (mudel 1), ja *alla* viitab olukorrale, kus inimese individuaalne riskihinnang langeb uut mudelit kasutades, võrreldes vana mudeliga. Seega $O(\text{üles}/\text{tulem})$ on sündmuste osakaal, kus individuaalne riskihinnang uut mudelit kasutades tõuseb tulemi saanud inimeste seas, ja $O(\text{alla}/\text{tulem})$ on sündmuste osakaal, kus riskihinnang uut mudelit kasutades langeb tulemi saanud inimeste seas. Samal meetodil arvutatakse välja ka hinnangud tulemita

inimeste kohta. Käesoleva töö raames oli *tulem* defineeritud kui inimesed, kes said jälgimisaja jooksul SVH, ja *tulemita* defineeritud kui inimesed, kes ei saanud jälgimisaja jooksul SVHd.

NRI test põhineb z -testil, mis järgib normaaljaotust ja eeldab sõltumatust tulemiga ja tulemita isikute vahel (108). Testi tulemusena saadavad p -väärtused on kahepoolsed. Reklassifitseerimise muutuse arvutamiseks kasutati R-i paketti *PredictionComparison* (109).

5. Tulemused

5.1 Tunnused uuringu algushetkel

5aastase jälgimisaja jooksul vaadeldi kokku 52 172 inimest, kellest 3,7% ($n = 1935$) said esmase SVH (Tabel 3). Kõikide uuritavate keskmine vanus oli 54,6 eluaastat (standardhälve SD 9,5). Tulemiga inimeste seas oli keskmine vanus 62,1 (SD 8,5) eluaastat, tulemita inimeste seas 54,4 (SD 9,4) eluaastat. Mõlema mudeli keskmine jälgimisaja pikkus treeningandmetel oli 1769 päeva ja testandmetel 1768 päeva. SVH saanutest 60,1% ($n = 1163$) olid mehed, tulemita inimeste seas oli mehi 45,0% ($n = 22 591$).

Kõige levinum kaasuv haigus kõikide uuritavate seas oli hüpertensioon (35,6%), sellele järgnesid ravimite tarvitamised – kortikosteroidide tarvitamine (12,0%), uinutite ja rahustite tarvitamine (10,8%). Võrreldes tulemita inimestega esines tulemiga inimeste seas oluliselt rohkem kaasuvaid haigusi ja ravimite tarvitamist. Neist kõige levinumad olid hüpertensioon (59,3%), 2. tüüpi diabeet (14,9%) ning uinutite ja rahustite tarvitamine (14,5%).

Tabel 3. 40–74aastaste uuringusse kaasatud inimeste tunnused jälgimisaja alguses

Tunnus	Tulemiga		Tulemita		Kõik uuritavad		<i>p</i> -väärtus*
	inimesed		inimesed		(n = 52 172)		
	(n = 1935)		(n = 50 237)				
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	
Keskmine vanus aastates (SD)**	62,1	(8,5)	54,4	(9,4)	54,6	(9,5)	< 0,001
Sugu = mees	1163	60,1	22 591	45,0	23 754	46,0	< 0,001
Hüpertensioon	1148	59,3	17 271	34,4	18 419	35,6	< 0,001
Kodade virvendus ja laperdus	173	8,9	1310	2,6	1483	3,0	< 0,001
Krooniline neeruhaigus	12	0,6	94	0,2	106	0,2	< 0,001
Migreen	17	0,9	922	1,8	939	1,8	0,003
Raske vaimne haigus	37	1,9	802	1,6	839	1,6	0,322
Reumatoidartriit	60	3,1	933	1,9	993	1,9	< 0,001
1. tüüpi diabeet	23	1,2	257	0,5	280	0,5	< 0,001
2. tüüpi diabeet	288	14,9	3097	6,2	3385	6,6	< 0,001
Antidepressantide tarvitamine	112	5,8	2950	5,9	3062	5,9	0,916
Kortikosteroidide tarvitamine	271	14,0	6012	12,0	6283	12,0	0,008
Uinutite ja rahustite tarvitamine	281	14,5	5366	10,7	5647	10,8	< 0,001

Paksus kirjas on välja toodud statistiliselt olulised seosed ($p < 0,05$).

**P*-väärtus näitab erinevust tulemiga ja tulemita inimeste valimite vahel.

**Vanusele vastavas tabeli reas on protsendi (%) asemel esitatud standardhälve (SD)

5.2 Südame-veresoonkonna haigustesse esmashaigestumine ja suremus

60,1% ($n = 1163$) SVH esmasjuhtudest tekkis meestel ja 39,9% ($n = 772$) naistel (Tabel 4). Kõige madalam haigestumine oli nii meeste kui ka naiste puhul 40–44aastaste vanuserühmas (3,8%), moodustades meeste seas 4,1% ja naiste seas 3,4% juhtumitest. Kõige kõrgem SVH esmashaigestumine oli 70–74aastaste vanuserühmas (22,9%), moodustades meeste seas 18,7% ja naiste seas 29,3% juhtumitest.

Vanuserühmades 50–54, 55–59 ja 65–69 esines meeste seas oluliselt rohkem SVH haigestumist võrreldes naistega (p -väärtused $< 0,05$). Vanuserühmas 70–74 esines naiste seas oluliselt rohkem SVH haigestumist võrreldes meestega ($p < 0,001$).

Tabel 4. SVH esmasjuhtude esinemine vanusrühma ja soo järgi

Tunnus	Naised (n = 772)		Mehed (n = 1163)		Kokku (n = 1935)		p-väärtus*
	n	%	n	%	n	%	
Vanuserühm							
40–44	26	3,4	48	4,1	74	3,8	0,468
45–49	34	4,4	94	8,1	128	6,6	0,998
50–54	39	5,1	127	11,0	166	8,6	< 0,001
55–59	102	13,2	204	17,5	306	15,8	0,013
60–64	159	20,5	241	20,7	400	20,7	0,948
65–69	186	24,1	231	19,9	417	21,6	0,030
70–74	226	29,3	218	18,7	444	22,9	< 0,001

Paksus kirjas on välja toodud statistiliselt olulised seosed ($p < 0,05$).

*P-väärtus näitab erinevust meeste ja naiste valimite vahel.

Jälgimisaja jooksul suri esmase SVH saanute ehk 1935 inimese seas vähemalt 399 inimest (20,6%), kellest 156 inimest (39,1%) surid andmete järgi samal päeval esmase SVH diagnoosi saamisega (Tabel 5). Ligi pooled surmadest (51,4%) esinesid esimese 30 päeva jooksul pärast esmast SVH diagnoosi. Rohkem kui 90 päeva pärast esmast SVH diagnoosi suri 163 inimest (40,9%).

Tabel 5. Esmase SVH järgselt surnud uuritavate jagunemine vastavalt surma ajale

Aeg	Surmad esmase SVH diagnoosi saanud inimeste seas (n = 399)	
	n	%*
Samal päeval SVH diagnoosiga	156	39,1
Esimese kuu jooksul	205	51,4
Esimese kahe kuu jooksul	224	56,1
Esimese kolme kuu jooksul	236	59,1
Rohkem kui kolm kuud pärast diagnoosi	163	40,9

*Osakaalud on arvatud SVHsse haigestunute põhjal.

Esmase SVH saanud naistest suri diagnoosi saamise järgselt 141 inimest (18,3%) ja meestest 258 inimest (22,2%) (Tabel 6). Kõige rohkem naisi ja mehi suri vanuserühmas 70–74 (36,6%), moodustades meeste seas 33,3% ja naiste seas 42,6% juhtumitest. Kõige madalam suremus oli 40–44aastaste seas (1,0%), moodustades naiste seas 0,7% ja meeste seas 1,2% juhtumitest.

Tabel 6. Esmase SVH järgselt surnud uuritavate surmad vanuserühma ja soo kaupa

Tunnus	Naised (<i>n</i> = 141)		Mehed (<i>n</i> = 258)		Suremus kokku (<i>n</i> = 399)		<i>p</i> -väärtus*
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	
Vanuserühm							
40–44	1	0,7	3	1,2	4	1,0	1
45–49	1	0,7	7	2,7	8	2,0	0,269
50–54	3	2,1	17	6,6	20	5,0	0,056
55–59	20	14,2	37	14,3	57	14,3	1
60–64	23	16,3	50	19,4	73	18,3	0,500
65–69	33	23,4	58	22,5	91	22,8	0,901
70–74	60	42,6	86	33,3	146	36,6	0,082

**P*-väärtus näitab erinevust meeste ja naiste vahel.

5.3 Mudelite tulemused

5.3.1 Tegurite ennustusvõime

Tulemused on raporteeritud sisemise valideerimise järgselt. Lõplikust mudelist jäid Lasso meetodi kasutamisel välja tunnused migreen ja kortikosteroidide tarvitamine. Nii vanus-soo mudelis (mudel 1) kui ka lõplikus mudelis (mudel 2) omas kõige tugevamat ennustusvõimet meessugu (Tabel 7). Riskimäärade suhted meessoole olid vastavalt mudelis 1 HR = 2,23 (95% CI 2,01–2,48) ja mudelis 2 HR = 2,30 (95% CI 2,06–2,56). Lõplikus mudelis olid lisaks meessoole kõige suurema ennustusvõimega hüpertensiooni (HR = 1,54, 95% CI 1,37–1,73), kodade virvenduse ja laperduse (HR = 1,49, 95% CI 1,22–1,78), reumatoidartriidi (HR = 1,41, 95% CI 1,02–1,89) ja 2. tüüpi diabeedi (HR = 1,41, 95% CI 1,21–1,64) diagnoosid. Koefitsiendid, mille põhjal tehti riskimäärade suhete arvutused on välja toodud Lisas 3. Lõpliku mudeli koefitsientide alusel tehtud näidiskalkulatsioonid riski hindamiseks kahe stsenaariumi näitel on välja toodud Lisas 4.

Tabel 7. Kahe mudeli tulemused tunnuste kaupa

Tunnus	Mudel 1	Mudel 2
	Vanus-soo mudel	Lõplik mudel
	HR (95% CI)	HR (95% CI)
Demograafilised tunnused		
Mees	2,23 (2,01–2,48)	2,30 (2,06–2,56)
Naine	1,00	1,00
Vanus (ühe aasta kaupa)	1,09 (1,09–1,10)	1,08 (1,08–1,09)
Diagnoosid		
Hüpertensioon		1,54 (1,37–1,73)
Kodade virvendus ja laperdus		1,49 (1,22–1,78)
Krooniline neeruhaigus		1,06 (0,80–2,12)
Raske vaimne haigus		1,01 (0,82–1,51)
Reumatoidartriit		1,41 (1,02–1,89)
1. tüüpi diabeet		1,11 (0,83–1,82)
2. tüüpi diabeet		1,41 (1,21–1,64)
Ravimite tarvitamine		
Antidepressandid		1,09 (0,90–1,36)
Uinutid ja rahustid		1,09 (0,95–1,26)

Paksus kirjas on välja toodud statistiliselt olulised seosed.

5.3.2 Diskrimineerimine

Vanus-soo mudeli (mudel 1) C-statistik treeningandmetel oli 0,750 (95% CI 0,738–0,761) ning testandmetel 0,747 (95% CI 0,727–0,768) (Tabel 8). Lõpliku mudeli (mudel 2) C-statistik treeningandmetel oli 0,762 (95% CI 0,751–0,773) ning testandmetel 0,757 (95% CI 0,737–0,778). Lõpliku mudeli C-statistik muutus võrreldes vanus-soo mudeliga +0,012 ühiku võrra treeningandmetel ja +0,010 ühiku võrra testandmetel.

Tabel 8. Kahe mudeli diskrimineerimine mõõdetuna Harrelli C-statistikuga

	Mudel 1	Mudel 2
	Vanus-soo mudel	Lõplik mudel
Treeningandmed		
C-statistik (95% CI)	0,750 (0,738–0,761)	0,762 (0,751–0,773)
C-statistiku muutus		+0,012
Testandmed		
C-statistik (95% CI)	0,747 (0,727–0,768)	0,757 (0,737–0,778)
C-statistiku muutus		+0,010

5.3.3 Reklassifitseerimise muutus

Tulemiga inimesed. Mõlemad mudelid ennustasid, et SVH risk on madal (ehk 5aastane risk saada SVH < 5%) 112 inimesel (Tabel 9). 3 inimest, kes vanus-soo mudeli alusel said madala riskihinnangu, said lõplikku mudelit kasutades kõrge riskihinnangu. Seega võrreldes vanus-soo mudeliga sai lõpliku mudeli alusel 0,2% rohkem inimesi kõrge riskihinnangu. 149 inimest, kes said vanus-soo mudeli alusel kõrge riskihinnangu, said lõpliku mudeli alusel madala riskihinnangu. Mõlemad mudelid ennustasid kõrge riskiga inimeste hulka 1671 inimest. Lõpliku mudeli alusel langes 7,7% inimeste riskihinnang madalamaks võrreldes vanus-soo mudeliga. Kokkuvõttes langes lõpliku mudeli kasutamisel õigesti klassifitseeritute hulk 7,5% ($p < 0,001$). See tähendab, et tulemiga inimeste seas langes lõpliku mudeli alusel ennustustäpsus 7,5% võrreldes vanus-soo mudeliga ehk inimesed, kes oleks pidanud saama kõrge riskihinnangu, said lõpliku mudeli alusel hoopis vale ehk madala riskihinnangu.

Tulemita inimesed. Mõlemad mudelid ennustasid, et SVH risk on madal (ehk 5aastane risk saada SVH < 5%) 15 490 inimesel. 138 inimest, kes vanus-soo mudeli alusel said madala riskihinnangu, said lõplikku mudelit kasutades kõrge riskihinnangu. Ehk võrreldes vanus-soo mudelit lõpliku mudeliga, sai viimast kasutades 0,3% rohkem inimesi madala riskihinnangu asemel kõrge riskihinnangu. Mõlemad mudelid ennustasid kõrge riskiga inimeste hulka 25 065 inimest. 9544 inimest, kes vanus-soo mudeli alusel said kõrge riskihinnangu, said lõplikku mudelit kasutades madala riskihinnangu. Seega võrreldes vanus-soo mudeliga, sai lõpliku mudeli alusel 19,0% rohkem inimesi madala riskihinnangu. Võrreldes vanus-soo mudeliga, klassifitseeriti lõpliku mudeli poolt õigesti ehk madalasse riskikategooriasse 18,7% rohkem inimesi ($p < 0,001$).

See tähendab, et tulemita inimeste seas tõusis lõpliku mudeli ennustustäpsus ning lõplik mudel eristas tulemita inimeste seas oluliselt paremini madala riskiga inimesi kõrge riskiga inimestest võrreldes vanus-soo mudeliga.

Võrreldes vanus-soo mudelit lõpliku mudeliga, oli uue mudeli tulemusel reklassifitseerimise muutus NRI 0,112 (95% CI 0,098–0,123). Positiivne NRI näitab, et lõplik mudel klassifitseerib inimesi kõrge ja madala riskihinnangu gruppidesse täpsemalt kui vanus-soo mudel, arvestades nii tulemiga kui tulemita inimeste gruppe.

Tabel 9. Reklassifitseerimise muutused kahe mudeli vahel

Tunnus	Mudel 1 Vanus-soo mudel	Mudel 2 Lõplik mudel		Reklassifitseeritud		Õigesti reklassifitseeritud	<i>p</i> -väärtus
		< 5%	> 5%	Tõusnud risk	Langenud risk		
Tulemiga inimesed (<i>n</i> = 1935)							
	< 5%	112	3	0,2%		-7,5%	< 0,001
	> 5%	149	1671		7,7%		
Tulemita inimesed (<i>n</i> = 50 237)							
	< 5%	15 490	138	0,3%		18,7%	< 0,001
	> 5%	9544	25 065		19,0%		
Reklassifitseerimise muutus NRI (95% CI)						0,112 (0,098–0,123)*	< 0,001

Paksus kirjas on välja toodud statistiliselt olulised seosed ($p < 0,05$). Helepunane taust näitab, kui lõplik mudel klassifitseeris inimesi riskigruppidesse ebatäpsemalt kui vanus-soo mudel. Roheline taust näitab, kui lõplik mudel klassifitseeris inimesi riskigruppidesse täpsemini kui vanus-soo mudel. Must-valge trüki puhul vastab tumedam toon rohelisele ja heledam toon punasele.

*Positiivne NRI näitab, et lõplik mudel klassifitseeris täpsemini kui vanus-soo mudel.

6. Arutelu

Käesoleva magistritöö eesmärk oli luua kaks riskimudelit, et ennustada esmase SVH tekkeriski 40–74aastaste tervete inimeste seas. Esimene mudel kaasas teguritena vanust ja sugu ning teine mudel kaasas teguritena lisaks eelmainitutele veel diagnooside ja ravimite tarvitamise infot. Töö tulemusena selgus, et lisategurite kaasamine parandas lõpliku mudeli diskrimineerimist võrreldes algse vanus-soo mudeliga. Kõige rohkem tõstsid lõplikus mudelis SVH riski hüpertensiooni, kodade virvenduse ja laperduse ning reumatoidartriidi diagnooside olemasolu.

Südame-veresoonkonna haigustesse esmashaigestumine ja surmad

Ajavahemikul 01.01.2015–31.12.2019 tuvastati 1935 SVH esmasjuhtu (3,7%). Rohkem kui pooled SVH esmasjuhtudest esinesid meeste seas. See on kooskõlas varasemate sarnaste uuringutega (76, 78). Vanuserühmades 50–54, 55–59 ja 65–69 esines meeste seas oluliselt rohkem SVH haigestumist võrreldes naistega ($p < 0,05$), kuid vanuserühmas 70–74 esines naiste seas oluliselt rohkem SVH haigestumist võrreldes meestega ($p < 0,001$). Selline tulemus leiti ka QRISK3 meeskonna poolt, kus alates 70. eluaastast oli naiste esmashaigestumine kõrgem kui meestel (78). Üks võimalik selgitus sellele tulemusele on see, et mehed saavad esmase SVH tunduvalt varem kui naised, mistõttu hilisemas eas esmashaigestub rohkem naisi (110, 111). Ka Tervise Arengu Instituudi Tervisestatistika andmebaasi alusel on sõltuvalt aastast esinenud naiste seas alates vanusest 65 kõrgemat südame isheemiatõve (I20–25) ja peaajuveresoonte haiguste (I60–69) esmashaigestumist võrreldes meestega (21).

Ligikaudu viiendik esmase SVH saanutest suri, kellest kolmandik suri samal päeval esmase SVH diagnoosi saamisega. Magistritöö aluseks olevast andmebaasist on puudu Surmaregistri andmed, mille tõttu on olemas umbes 67% surmadest, mis tähendab, et tõenäoliselt on suremus antud andmebaasis oluliselt alahinnatud. Ka varasemates uuringutes on selgunud, et Surmaregistri andmete mitte kaasamine võib sõltuvalt andmebaasi eripäradest viia suremuse üle- või alahindamiseni (93). Samuti pole võimalik täie kindlusega väita, et vaadeldud surmad olid põhjustatud SVHst.

Mudelisse kaasatud tegurite levimus ja ennustusvõime

Riskitegurite levimus varieerus 40–74aastaste inimeste seas jälgimisaja alguses sõltuvalt tegurist 0,2–35,6%ni. Tulemiga inimeste seas esines võrreldes tulemita inimestega oluliselt rohkem kaasuvaid haigusi ja ravimite tarvitamist pea kõikide mudelisse kaasatud tegurite osas. Olulisi

erinevusi ei esinenud raske vaimse haiguse diagnoosi olemasolu ja antidepressantide tarvitamise osas.

Võrreldes teiste uuringutega, olid antud töö raames riskitegurite levimused sageli kõrgemad (76, 78). Riikide ja andmebaaside vahelised erisused võivad tekkida mitmel põhjusel. Esiteks võib erinevusi põhjustada see, kuidas ja kui sageli inimesed uuritavas riigis arsti juures käivad. Samas ei pruugi kõrgema levimuse taga alati olla seosed arsti juures käimisega. Mõjutada võib ka see, kui palju on igas riigis teatud haigusega inimesi ja millised on kohalikud diagnoosimise praktikad. Näiteks ei ole Eestis tavaks kroonilise neeruhaiguse puhul eristada RHK-10 põhiseid raskusastmeid (koodid N18.3–18.5). Seetõttu tuli magistrیتöösse kaasata inimesed vastavalt eGFR mõõtmistulemusele, mis vastasid raskusastmetele 3–5. Samas QRISK3 raames eristati haiguse raskusastmeid diagnoosikoodide põhiseilt (78). Erinevuste taga võib olla ka see, millised on diagnoosimiseks kasutatavad klassifitseerimise süsteemide eripärad. Eestis on kasutusel RHK-10 süsteem, kuid QRISK3 loomise hetkel kasutati perearstide seas READ-koodide süsteemi (78), mis võib oma detailsuselt ja kasutamisevõimaluste poolest tuua erisusi haiguse diagnoosimisel.

Lisategurite kaasamine parandas lõpliku mudeli täpsust võrreldes mudeliga, kuhu oli kaasatud ainult vanus ja sugu. Mudelisse kaasatud tegurid olid olulised SVH riski ennustamisel, kuid võrreldes varasemate mudelitega olid riskimäärade suhted enamjaolt madalamad (78). Lasso meetodil eemaldati migreeni diagnoos, mis oli ebaoluline SVH riski ennustamisel. Diagnooside grupist eristus reumatoidartriidi diagnoos, mille riskimäärade suhe (HR = 1,41; 95% CI 1,02–1,89) oli kõrgem võrreldes QRISK3 naiste (HR = 1,24; 95% CI 1,20–1,27) ja meeste (HR = 1,23; 95% CI 1,19–1,28) mudeliga (78). Kuigi QRISK3 mudelis oli kortikosteroidide tarvitamine oluline tegur nii meeste (HR = 1,58; 95% CI 1,50–1,66) kui naiste mudelis (HR = 1,81; 95% CI 1,74–1,89) (78), siis käesoleva töö raames loodud mudelis osutus kortikosteroidide tarvitamine ebaoluliseks ja see tegur eemaldati mudelist Lasso meetodil.

Erisused mudeli tulemustes võrreldes varasemate mudelitega võivad tuleneda sellest, et mudelid on loodud erinevate rahvastike põhjal ning varem loodud mudelid on kaasanud kordades rohkem uuritavaid. Kui magistrیتöö raames treeniti mudel, kus oli 1935 SVH juhtumit, siis QRISK3 naiste mudel treeniti 160 549 ja meeste mudel 203 016 SVH juhtumi põhjal (78) ning SCORE2 mudel treeniti 30 121 juhtumi põhjal (76). Tugeva ennustusvõime tuvastamiseks peab olema piisavalt suur valim ja statistiline võimsus (13). Seega väiksem valim ja madalam sündmuste arv (nii SVH kui mudelisse kaasatud tegurite põhjal) võib viia selleni, et mudelisse kaasatud tegurid

on väiksema ennustusvõimega. Samas töö aluseks olevas andmebaasis oli iga lõplikkusse kaasatud teguri (kokku 13 tegurit) kohta keskmiselt 149 SVH juhtu, mis võiks teoreetiliselt olla piisav (13).

Lisaks valimi suurusele ja juhtumite arvule võib mudeli tulemusi mõjutada see, kuidas tunnused on mudelisse defineeritud ning diagnoosimise erisused. Näiteks võib riigiti erineda diagnoosimise detailsus, kodeerimine ning diagnooside üle- või aladiagnoosimine. Samuti võisid antud töö raames olla ravimite tarvitamise kriteeriumid liiga ranged. Näiteks QRISK3 kaasas ravimite välja kirjutamise infot (78), kuid käesolevas töös kasutati ravimite välja ostmise andmeid. Osad diagnoosid võisid olla defineeritud liiga laialt. Näiteks lähtuti käesolevas töös hüpertensiooni defineerimisel laiemast definitsioonist, samas QRISK3 kasutas hüpertensiooni diagnoosi, mille puhul läksid arvesse ainult juhud, kus haigust ka raviti, mis on tunduvalt kitsam (78). Kitsam defineerimine ja tegurite madalam levimus võivad viia suuremate riskimäärade suheteni.

Loodud mudelite diskrimineerimisvõime ja reklassifitseerimise muutus

Mõlema mudeli usaldusvahemikud nii treening- kui testandmetel ületasid aktsepteeritava taseme alampiiri 0,70 ehk mõlemad mudelid on lähtuvalt Harrelli tõlgendusest aktsepteeritavad (13). Mudelite diskrimineerimine erines treening- ja testandmetel vähesel määral, samuti olid kõik usaldusvahemikud kitsad, mis viitab, et mudelid ei ole tõenäoliselt üle treenitud ning need võiksid sarnastel tingimustel kogutud tundmatutel andmetel anda küllaltki stabiilseid tulemusi (13).

Mõlema loodud mudeli usaldusvahemikud kattusid omavahel, aga nende omavaheline võrdlemine ei ole C-statistiku looja, F. Harrelli poolt soovitatud, sest see ei võimalda kinnitada seose olemasolu või selle puudumist (13, 112). Võrreldes vanus-soo mudeliga tõusis lõpliku mudeli C-statistik treeningandmetel +0,012 ja testandmetel +0,010 ühiku võrra. On keeruline hinnata, kas lõpliku mudeli C-statistiku tõus on oluline, kuna juhistes teadlastele pole kindlaid piire paika pandud (72). Üldiselt käsitletakse tõusu $< 0,005$ marginaalsena (113). Samas tõusu vahemikus 0,006–0,015 peetakse rahuldavaks (114). Tõusu vahemikus 0,020–0,030 ühikut peetakse väga heaks, kuid seda juhtub harva. Näiteks võrreldes QRISKi algset mudelit QRISK2 mudeliga, tõusis lisategurite abil C-statistik 0,002 ühiku võrra (100). Võrreldes QRISK2 mudelit QRISK3 mudeliga, paranes diskrimineerimine 0,001 ühiku võrra (78). Seega QRISK2 ja QRISK3 mudelite vahel ei olnud suurt diskrimineerimise muutust. Sellest hoolimata on Inglismaa ravijuhiste järgi soovitatud kasutada QRISK3 mudelit (115), kuna mudeli rakendamiseks ei pea mudeli diskrimineerimine paranema suurel määral ning rolli mängivad ka muud mudeli näitajad. Magistritöö raames loodud lõpliku mudeli C-statistiku tõus on lähtuvalt varasemate uuringute

tõlgendustest rahuldav, kuid kliinilises mõttes kindlasti oluline, sest see tähendab mudeli täpsuse paranemist.

Töö raames loodud mudelite diskrimineerimine jääb alla eelmisele SCORE mudelile, mis on andnud Eesti Geenivaramu andmetel väga häid tulemusi. Näiteks Saar *et al* (2019) uuringu alusel oli Eesti Geenivaramu põhjal testitud SCORE mudeli C-statistik 0,865 (83) ning Tillmann *et al* (2020) uuringu alusel oli SCORE C-statistik enne kalibreerimist 0,832 (95% CI 0,769–0,896) ja pärast kalibreerimist 0,851 (95% CI 0,791–0,910) (57). Samas oli nii magistritöö raames loodud lõpliku mudeli kui ka vanus-soo mudeli diskrimineerimine oluliselt parem kui SCORE2 mudelil, mis oli Eesti Geenivaramu andmetel 0,694 (95% CI 0,650–0,739) (76). Lähtuvalt Harrelli tõlgendusest on SCORE2 Eesti andmetel andnud ebaadekvaatseid tulemusi (13). Seega SCORE varasem versioon töötas Eesti andmetel mitme uuringu põhjal väga hästi, kuid SCORE2 tulemused langesid olulisel määral. On raske öelda, millest võib olla tingitud SCORE2 diskrimineerimise väga suur langus võrreldes SCORE mudeliga. Üks põhjus sellise erinevuse taga võib olla see, et Geenivaramu andmebaasi kooslus võib erineda üldpopulatsioonist. Teine põhjus võib olla see, et mudeleid on testitud väikestel valimitel. Olenemata sellest, millist mudelit otsustatakse rakendada, võiks Harrelli C-statistik olla vähemalt aktsepteeritaval tasemel (C-statistik > 0,70) (13). ESC soovib praeguste juhiste kohaselt SVH riski hindamiseks kasutada just SCORE2 mudelit (8), kuid arvestades selle mudeli ebaadekvaatset diskrimineerimist Eesti Geenivaramu andmetel, vajaks SCORE2 enne laiemat kasutust rek calibreerimist Eesti andmetele. Enne, kui seda tehakse, peaks SCORE2 kasutamisel olema ettevaatlik. Magistritöö raames loodud mudelid on juba treenitud ja testitud Eesti andmetel, mistõttu võivad need anda võrreldes rek calibreerimata SCORE2 mudeliga täpsemaid tulemusi.

NRI tulemus 0,112 (95% CI 0,098–0,123) viitab sellele, et magistritöö raames loodud lõplik mudel on inimeste klassifitseerimises riskigruppidesse täpsem kui vanus-soo mudel. Parema klassifitseerimine võimaldab tuvastada potentsiaalselt rohkem inimesi, keda peaks kutsuma edasistele uuringutele ja kes vajavad edasisi sekkumisi, sh statiinravi. Võrreldes vanus-soo mudeliga oli lõplik mudel 7,5% halvem tulemiga inimeste klassifitseerimises ja 18,7% parem tulemita inimeste klassifitseerimises. Seega haigestunute seas langes sensitiivsus, mittejuhtumite puhul tõusis spetsiifilisus.

Reklassifitseerimise muutus NRI võib olla oma olemuselt ebaoluline ja eksitav number. Tulemi ja tulemita inimeste hinnangud on üsna lihtsasti tõlgendatavad, kuid nende kahe summa ehk NRI aluseks on eeldus, et valesti klassifitseerimise kahju on võrdne nii tulemiga kui tulemita

inimeste seas (116). Teisisõnu, NRI eeldus on, et kui kahe mudeli vahel valenegatiivsete hulk langeb, siis on see sama tähtis kui sama suur valepositiivsete tõus. Sellise eelduse korral tuleks sekkumist rakendada siis, kui inimese riskimäär on $> 50\%$. Tegelikult soovivad kliinilised juhised sekkumisi rakendada kordades madalamate riskimääradega. Näiteks statiinravimeid soovitatakse kõige hiljutisema soovitus järgi kasutada siis, kui inimese 10aastane risk saada SVH on 10% või kõrgem (8). Seega antud kliinilises küsimuses on valenegatiivsete vältimine kordades olulisem eesmärk kui valepositiivsete vältimine. Soovituslik on kasutada kaalutud NRId, kus mõlemale poolele on lisatud oma kaal (117). Arvutuste kohaselt said magistritöö raames loodud lõplikku mudelit kasutades kasu hinnanguliselt 9260 inimest (17,7%). See muutus toimus peamiselt inimeste seas, kellel on madal risk ehk kes ei saanud tulemit. Selleks, et otsustada, kas mudel on kasulik ka siis, kui tulemiga ja tulemita inimeste kaalud pole võrdsed, tuleks teha täiendavaid analüüse, nagu näiteks otsustuskõvera analüüs (*decision curve analysis*) (118, 119).

Magistritöö puudused ja tugevused

Magistritöös esines mitmeid puudusi. Valim oli võrreldes teiste SVH mudelite aluseks kasutatavate valimitega tunduvalt väiksem ning seetõttu ei loodud soospetsiifilisi mudeleid. Meeste ja naiste puhul võivad mudelisse kaasatavad tegurid omada erinevat ennustusvõimet. Sellest tulenevalt võivad riskid antud töös olla ala- või ülehinnatud. See võib mõjutada ka mudeli tulemusi olukorras, kus mudelit rakendatakse uuel andmestikul.

Töös kasutatud andmebaas oli piiratud 8aastase ajaperioodiga, mistõttu oli inimesi võimalik jälgida tavapärase 10aastase jälgimisaja asemel 5 aastat. See võib mõjutada mudeli tulemusi, sest osadel inimestel, kellel võiks 10aastase perioodi jooksul tekkida esmane SVH, seda 5aastase perioodi jooksul välja ei tulnud. Lisaks puudusid uuringu tegemiseks kasutatud andmebaasist Rahvastikuregistri ja Surmaregistri andmed, mistõttu ei saanud eristada inimesi, kes uuringuperioodil riigist lahkusid, ning fataalsete juhtumite usaldusväärsus oli madal, mille tõttu ei saanud hinnata fataalse lõppega SVH riski. Sellest tulenevalt võib loodud mudel SVH riski alahinnata.

Töösse kaasatud tegurid võisid olla defineeritud liiga laialt või kitsalt. Lisaks oli krooniline neeruhaigus võimalik defineerida ainult läbi mõõtmistulemuste, kuid inimese risk võib sõltuvalt mõõtmistulemusest olla väga erinev. On suur vahe, kas inimene langeb kroonilise neeruhaiguse 3. või 5. raskusastmesse. Kuna Eestis raskusastme põhiseid diagnoose riskitegurite info kogumise

perioodil diagnoosides ei kajastunud, oleks täpsem olnud kroonilise neeruhaiguse tunnus kaasata lähtuvalt mõõtmistulemusest pideva, mitte binaarse tunnusena.

Valitud pidevate tunnuste puuduvus ajaperioodil 01.01.2012–31.12.2014 oli üle 50%, mistõttu ei olnud võimalik mudelisse kaasata mitmeid pidevaid tunnuseid ja hinnata nende eraldiseisvat ennustusvõimet SVH tekkes. Selline andmete puuduvus võib tuleneda vähesest dokumentide esitamisest kesksesse andmekogudesse ja/või mõõtmistulemuste raporteerimisest, mis võis olla sellel perioodil vähem levinud praktika. Näiteks oli üldkolesterooli mõõtmine tehtud 52% le uuritavatest, kuid mõõtmistulemus oli andmetes olemas 26%l. Samuti on töö puuduseks see, et mudelisse ei kaasatud suitsetamise staatust ning erinevaid interaktsioone tegurite vahel, mida varasemad mudelid on teinud. Senisest enam tähelepanu tuleks pöörata mõõtmistulemuste ja suitsetamise staatuse info tuvastamisele vabatekstist, mis võib aidata puuduvate tunnuste osakaalu sellel ajaperioodil vähendada.

Magistritöö raames hinnati kahe mudeli klassifitseerimist ning selle alusel oli lõplik mudel inimeste klassifitseerimises riskikategooriatesse parem. Küll aga ei ole teada, kas selline tulemus kehtib ka siis, kui eeldada, et tulemi saanud ja tulemita inimesed ei anna sellele hinnangule võrdset kaalu. Töö valimist ei eemaldatud inimesi, kes võisid juba kasutada SVHd ennetavaid ravimeid (näiteks statiinravimid), mistõttu võis see viia näiliselt parema diskrimineerimiseni.

Magistritööl on mitmeid tugevusi. Magistritöö tegemiseks kasutatud andmebaasi loomisel kasutati kolme riiklikku andmekogu, mis aitab kompenseerida ühe või teise andmekogu puuduvaid andmeid (näiteks diagnooside osas), võimaldades saada inimesest terviklikum pilt. Mudelite loomiseks kasutati elektroonilisi terviseandmeid, mistõttu oli võimalik kaasata kõik uuritavad. Kasutatud andmebaas põhines juhuvalimil, seetõttu on risk, et uuringusse kaasatud inimesed oleks tavapärasest olulisemalt haigemad või tervemad, väga madal. Mudel läbis 10kordse sisemise ristvalideerimise, mis tõstab mudeli usaldusväärsust ja langetab riski, et mudel on üle treenitud. Sellele viitavad ka küllaltki sarnased diskrimineerimise tulemused treening- ja testandmetel.

Kuigi maailma mõistes on varem loodud palju erinevaid SVH riskimudeleid, on vähesed neist tehtud elektrooniliste terviseandmete põhjal. Töö tulemusena selgus, et diagnoosid ja ravimite tarvitamise info on Eesti elektroonilistest terviseandmetest hästi tuvastatavad ning sellise info kaasamise tulemusena saab täpsema mudeli võrreldes vanus-soo mudeliga. Käesolev töö on ainulaadne, kuna keskenduti mudelite loomisele, mida on võimalik lisauuringute teostamisel paigutada SVH sõeluuringu varasemasse etappi, et sellega kindlaks määrata, keda potentsiaalselt kutsuda edasistele uuringutele.

7. Järeldused ja ettepanekud

Magistritöö raames uuriti SVH esmashaigestumist ja suremust, mudelisse kaasatud tegurite ennustusvõimet ja loodi kaks mudelit esmase SVH tekkeriski ennustamiseks. Töö tulemustel põhinevad järgnevad järeldused:

1. 5aastase jälgimisaja jooksul said esmase SVH 3,7% uuritavatest. Kõige levinumad riskitegurid uuritavate seas olid hüpertensiooni diagnoos, uinutite ja rahustite tarvitamine ning kortikosteroidide tarvitamine.
2. Lõpliku mudeli alusel oli meestel 2,32 korda kõrgem risk saada esmane SVH kui naistel. SVH riski suurendasid ka kaasuvad haigused ja ravimite tarvitamine. Hüpertensiooniga inimeste risk saada SVH oli 1,53 korda kõrgem ja kodade virvenduse või laperduse diagnoosiga inimeste risk saada SVH oli 1,48 korda kõrgem kui ilma nende diagnoosideta inimestel.
3. Mõlema mudeli diskrimineerimine testandmetel oli aktsepteeritav ja need võiksid sobida kasutamiseks Eestis. Diagnooside ja ravimite tarvitamise info kaasamine lõplikusse mudelisse parandas diskrimineerimist 0,012 ühiku võrra treening- ja 0,010 ühiku võrra testandmetel.
4. Lõpliku mudeli reklassifitseerimise muutus oli võrreldes vanus-soo mudeliga positiivne eeldusel, et tulemita ja tulemiga inimeste riski täpne klassifitseerimine on sama oluline.

Magistritööl põhinevad ettepanekud ning töö tulemuste rakendamine:

1. Pidevate tunnuste puuduvus ajaperioodil 2012–2014 oli üle 50%. See on oluline probleem, kui riskimudeli treenimiseks kasutatakse retrospektiivseid andmeid. Edasistes teadustöodes tuleks hinnata pidevate tunnuste imputeerimisvõimalusi või kaaluda uuringu riskitegurite kogumise perioodi hilisemaks lükkamist eeldusel, et ajas muutub mõõtmistulemuste edastamine elektrooniliste terviseandmete põhisesse andmebaasidesse ning andmete tuvastamine vabatekstist järjest paremaks.
2. Magistritöö raames loodud lõplik mudel vajab enne rakendamist täiendavaid analüüse. On vaja hinnata lõpliku mudeli tegelikku klassifitseerimist eeldusel, et tulemi saanud ja tulemita inimeste kaalud ei ole võrdsed. Mudeli diskrimineerimise ja klassifitseerimise parandamiseks võib kaaluda järgnevaid samme: lisategurite ja interaktsioonide kaasamine; tegurite definitsioonide konkreetsemaks muutmine; riskitegurite levimuse hindamine ja

võrdlus mõne teise andmebaasiga, et paremini mõista elektrooniliste terviseandmete eripärasid.

3. Mudeli prospektiivseks rakendamiseks võiks kaaluda mudeli baasriski kohandamist, kuna mudel on treenitud retrospektiivselt ning baasrisk võib arvestades SVH trende ajas langeda.
4. SVH riskihindamisest tasub mõelda kui mitmeetapilisest protsessist. Klassikaliselt kasutatakse esimeses etapis inimeste vanust ja sugu, et teha esmane riskihinnang, kuid esimene etapp võiks hõlmata ka olemas olevate andmete kaasamist elektroonilistest terviseandmetest (näiteks kaasuvad haigused, ravimite tarvitamine), et otsustada, kas inimene kuulub potentsiaalse sekkumise sihtrühma. Seda saab teha elektrooniliste terviseandmete põhjal loodud riskimudeli abil. Teine etapp võiks olla täiendavate andmete kogumine, kutsudes sihtrühma kuuluvad inimesed täiendavatele analüüsidele (näiteks vererõhu ja kolesterooli mõõtmised). Selline tegevuskava võiks potentsiaalselt olla mugavam, vastuvõetavam ja kulutõhusam lähenemine SVH sõeluuringuprogrammile. Näiteks kopsuvähi sõeluuringu teostatavusuuringus tuvastati uuritavate suitsetamise staatus kas perearsti infosüsteemi ehk terviseandmete põhjal, telefoni teel või mõne muu visiidi käigus (120). Seega kaheetapiline lähenemine sõeluuringule on võimalik, kuid enne mistahes sekkumise või tööriista rakendamist on vajalik analüüsida ja uurida kõiki sellega seotud eetilisi ja andmekaitsega seotud aspekte koostöös ekspertidega.

8. Kasutatud kirjandus

1. Roth GA, Mensah GA, Johnson CO, et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study. *J Am Coll Cardiol* 2020;76:2982–3021.
2. Magnussen C, Ojeda FM, Leong DP, et al. Global effect of modifiable risk factors on cardiovascular disease and mortality. *N Engl J Med* 2023;389:1273–85.
3. Timmis A, Vardas P, Townsend N, et al. European Society of Cardiology: Cardiovascular Disease Statistics 2021. *Eur Heart J* 2022;43:716–99.
4. World Health Organization. Cardiovascular diseases. (https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1). [10.01.2024].
5. McGill HC, McMahan CA, Gidding SS. Preventing heart disease in the 21st century. *Circulation* 2008;117:1216–27.
6. Rossello X, Dorresteijn JAN, Janssen A, et al. Risk prediction tools in cardiovascular disease prevention: A report from the ESC Prevention of CVD Programme led by the European Association of Preventive Cardiology (EAPC) in collaboration with the Acute Cardiovascular Care Association (ACCA) and the Association of Cardiovascular Nursing and Allied Professions (ACNAP). *EJCN* 2019;18:534–44.
7. Cooney MT, Selmer R, Lindman A, et al. Cardiovascular risk estimation in older persons: SCORE O.P. *Eur J Prev Cardiol* 2016;23:1093–103.
8. Visseren F, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice: Developed by the Task Force for cardiovascular disease prevention in clinical practice with representatives of the European Society of Cardiology and 12 medical societies. *Eur Heart J* 2021;42:3227–337.
9. World Health Organization. Implementation tools: package of essential noncommunicable (PEN) disease interventions for primary health care in low-resource settings. ([https://www.who.int/publications/i/item/implementation-tools-package-of-essential-noncommunicable-\(-pen\)-disease-interventions-for-primary-health-care-in-low-resource-settings](https://www.who.int/publications/i/item/implementation-tools-package-of-essential-noncommunicable-(-pen)-disease-interventions-for-primary-health-care-in-low-resource-settings)). [06.04.2024].
10. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;3:2011–30.

11. Booth HP, Prevost AT, Gulliford MC. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. *Pharmacoepidemiol Drug Saf* 2013;22:1357–61.
12. Payne RA, Abel GA, Simpson CR. A retrospective cohort study assessing patient characteristics and the incidence of cardiovascular disease using linked routine primary and secondary care data. *BMJ Open* 2012;2:e000723.
13. Steyerberg EW. Clinical prediction models. *Statistics for biology and health*. 2019.
14. Tervise infosüsteem. TEHIK. (<https://www.tehik.ee/tervise-infosusteem>). [07.03.2024].
15. Harron K. Data linkage in medical research. *BMJ Medicine* 2022;1:e000087.
16. RHK-10. Sotsiaalministeerium. (<https://rhk.sm.ee/>). [21.02.2024].
17. World Health Organization. Premature mortality from noncommunicable disease. (<https://www.who.int/data/gho/indicator-metadata-registry/imr-details/3411>). [07.02.2024].
18. Lababidi H, Salerno PR, Wass SY, et al. The global burden of premature cardiovascular disease, 1990–2019. *Int J Cardiol* 2023;19:e200212.
19. Eurostat. Causes of death statistics, 2023. (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Causes_of_death_statistics). [10.02.2024].
20. Statistikaamet. Statistika andmebaas. RV56: Aasta, surmapõhjuse RHK-10 järgi, sugu ning vanuserühm. (https://andmed.stat.ee/et/stat/rahvastik__rahvastikusundmused__surmad/RV56/table/tableViewLayout2). [10.02.2024].
21. Tervisestatistika ja terviseuuringute andmebaas. EH10: Esmashaigusjuhud soo ja vanuserühma järgi (1998–2016). (https://statistika.tai.ee/pxweb/et/Andmebaas/Andmebaas__02Haigestumus__01Esmashai gestumus/EH10.px/table/tableViewLayout2/). [08.04.2024].
22. Tervisestatistika ja terviseuuringute andmebaas. SD21: Surmad põhjuse, soo ja vanuserühma järgi. (https://statistika.tai.ee/pxweb/et/Andmebaas/Andmebaas__01Rahvastik__04Surmad/SD21.px/table/tableViewLayout2/). [08.04.2024].
23. Roth GA, Abate D, Abate KH, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet* 2018;392:1736–88.

24. British Heart Foundation. Reduce your risk of developing atherosclerosis. (<https://www.bhf.org.uk/information-support/heart-matters-magazine/medical/blocked-arteries/reduce-your-risk-factors>). [08.02.2024].
25. Tian F, Chen L, Qian Z (Min), et al. Ranking age-specific modifiable risk factors for cardiovascular disease and mortality: evidence from a population-based longitudinal study. *EClinicalMedicine* 2023;64.
26. Kannel WB, Doyle JT, Shephard RJ, et al. Prevention of cardiovascular disease in the elderly. *J Am Coll Cardiol* 1987;10(2 Suppl A):25A-28A.
27. Whelton PK, Carey RM, Aronow WS, et al. 2017 guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. *J Hypertens* 2018;71:1269–324.
28. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH guidelines for the management of arterial hypertension: the task force for the management of arterial hypertension of the European Society of Cardiology (ESC) and the European Society of Hypertension (ESH). *Eur Heart J* 2018;39:3021–104.
29. Müokardiinfarktregister. MIR aruanne 2021. (<https://www.kliinikum.ee/infarkt/aruanded/>). [10.02.2024].
30. Rapsomaniki E, Timmis A, George J, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1.25 million people. *Lancet* 2014;383:1899–911.
31. Lewington S, Clarke R, Qizilbash N, et al. Age-specific relevance of usual blood pressure to vascular mortality: A meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002;360:1903–13.
32. Singh GM, Danaei G, Farzadfar F, et al. The age-specific quantitative effects of metabolic risk factors on cardiovascular diseases and diabetes: a pooled analysis. *PLoS One* 2013;8.
33. Stevens SL, Wood S, Koshiaris C, et al. Blood pressure variability and cardiovascular disease: systematic review and meta-analysis. *The BMJ* 2016;354.
34. National Health Service. High cholesterol. Cholesterol levels, 2022. (<https://www.nhs.uk/conditions/high-cholesterol/cholesterol-levels/>). [10.02.2024].
35. American Heart Association. What is Cholesterol? (<https://www.heart.org/en/health-topics/cholesterol/about-cholesterol>). [10.02.2024].

36. World Health Organization. Noncommunicable diseases: risk factors. (<https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/ncd-risk-factors>). [10.02.2024].
37. Pirillo A, Casula M, Olmastroni E, et al. Global epidemiology of dyslipidaemias. *Nat Rev Cardiol* 2021;18:689–700.
38. Stamler J, Daviglius ML, Garside DB, et al. Relationship of baseline serum cholesterol levels in 3 large cohorts of younger men to long-term coronary, cardiovascular, and all-cause mortality and to longevity. *JAMA* 2000;284:311–8.
39. Jung E, Kong SY, Ro YS, et al. Serum cholesterol levels and risk of cardiovascular death: A systematic review and a dose-response meta-analysis of prospective cohort studies. *Int J Environ Res Public Health* 2022;19:8272.
40. Abdullah SM, Defina LF, Leonard D, et al. Long-term association of low-density lipoprotein cholesterol with cardiovascular mortality in individuals at low 10-year risk of atherosclerotic cardiovascular disease. *Circulation* 2018;138:2315–25.
41. Ference BA, Ginsberg HN, Graham I, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J* 2017;38:2459.
42. Law MR, Wald NJ, Thompson SG. By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? *BMJ* 1994;308:367.
43. World Health Organization. Health service delivery framework for prevention and management of obesity. (<https://www.who.int/publications/i/item/9789240073234>). [10.02.2024].
44. World Health Organization. Obesity. (https://www.who.int/health-topics/obesity#tab=tab_1). [10.02.2024].
45. La Sala L, Pontiroli AE. Prevention of diabetes and cardiovascular disease in obesity. *Int J Mol Sci* 2020;21:1–17.
46. Dwivedi AK, Dubey P, Cistola DP, et al. Association between obesity and cardiovascular outcomes: Updated evidence from meta-analysis studies. *Curr Cardiol Rep* 2020;22:25.

47. Sun YQ, Burgess S, Staley JR, et al. Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses. *BMJ* 2019;364:1042.
48. Wormser D, Kaptoge S, Di Angelantonio E, et al. Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: collaborative analysis of 58 prospective studies. *Lancet* 2011;377:1085–95.
49. Riaz H, Khan MS, Siddiqi TJ, et al. Association between obesity and cardiovascular outcomes: A systematic review and meta-analysis of mendelian randomization studies. *JAMA Netw Open* 2018;1:e183788.
50. Khan SS, Ning H, Wilkins JT, et al. Association of body mass index with lifetime risk of cardiovascular disease and compression of morbidity. *JAMA Cardiol* 2018;3:280–7.
51. World Health Organization. Tobacco. (https://www.who.int/health-topics/tobacco#tab=tab_1). [10.02.2024].
52. World Health Organization. WHO global report: mortality attributable to tobacco. (<https://www.who.int/publications/i/item/9789241564434>). [10.02.2024].
53. Doll R, Peto R, Boreham J, et al. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ* 2004;328:1519.
54. Prescott E, Hippe M, Schnohr P, et al. Smoking and risk of myocardial infarction in women and men: longitudinal population study. *BMJ* 1998;316:1043–7.
55. Wood D, Wray R, Poulter N, et al. JBS 2: Joint British Societies' guidelines on prevention of cardiovascular disease in clinical practice. *Heart* 2005;91(suppl 5).
56. Lv X, Sun J, Bi Y, et al. Risk of all-cause mortality and cardiovascular disease associated with secondhand smoke exposure: a systematic review and meta-analysis. *Int J Cardiol* 2015;199:106–15.
57. Tillmann T, Läll K, Dukes O, et al. Development and validation of two SCORE-based cardiovascular risk prediction models for Eastern Europe: a multicohort study. *Eur Heart J* 2020;41:3325–33.
58. Cook DG, Pocock SJ, Shaper AG, et al. Giving up smoking and the risk of heart attacks. A report from The British Regional Heart Study. *Lancet* 1986;2:1376–80.
59. Campbell NC, Thain J, Deans HG, et al. Secondary prevention in coronary heart disease: baseline survey of provision in general practice. *BMJ* 1998;316:1430–4.

60. Wald NJ, Watt HC. Prospective study of effect of switching from cigarettes to pipes or cigars on mortality from three smoking related diseases. *BMJ* 1997;314:1860–3.
61. World Health Organization. Diabetes. (https://www.who.int/health-topics/diabetes#tab=tab_1). [12.02.2024].
62. Uusitupa M, Khan TA, Vigiuliouk E, et al. Prevention of type 2 diabetes by lifestyle changes: a systematic review and meta-analysis. *J Nutr* 2019;11:2611.
63. Sarwar N, Gao P, Kondapally Seshasai SR, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet* 2010;375:2215–22.
64. Hajar R. Risk factors for coronary artery disease: historical perspectives. *Heart Views* 2017;18:109.
65. Matheus ASDM, Tannus LRM, Cobas RA, et al. Impact of diabetes on cardiovascular disease: an update. *Int J Hypertens* 2013;2013:15.
66. Jousilahti P, Laatikainen T, Peltonen M, et al. Primary prevention and risk factor reduction in coronary heart disease mortality among working aged men and women in eastern Finland over 40 years: population based observational study. *BMJ* 2016;352.
67. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA Guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* 2019;140:e596–646.
68. National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction, including lipid modification. Overview. (<https://www.nice.org.uk/guidance/ng238>). [19.02.2024].
69. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham Study. *Am J Cardiol* 1976;38:46–51.
70. Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med* 2004;66:411–421.
71. OHDSI. The book of OHDSI: Observational health data sciences and informatics. OHDSI, 2019.
72. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Med* 2015;13:1–10.

73. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 2016;69:245–7.
74. Reps JM, Schuemie MJ, Suchard MA, et al. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *JAMIA* 2018;25:969–75.
75. Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016;353:i2416.
76. SCORE2 working group, ESC cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J* 2021;42:2439–54.
77. SCORE2-OP working group, ESC cardiovascular risk collaboration. SCORE2-OP risk prediction algorithms: estimating incident cardiovascular event risk in older persons in four geographical risk regions. *Eur Heart J* 2021;42:2455–67.
78. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;357.
79. National Health Service. NHS Health Check. (<https://www.nhs.uk/conditions/nhs-health-check/>). [04.05.2024].
80. Goff DC, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA Guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* 2014;63:2935–59.
81. Hajifathalian K, Ueda P, Lu Y, et al. A novel risk score to predict cardiovascular disease risk in national populations (GloboRisk): a pooled analysis of prospective cohorts and health examination surveys. *Lancet Diabetes Endocrinol* 2015;3:339–55.
82. Farzadfar F. Cardiovascular disease risk prediction models: challenges and perspectives. *Lancet Glob Health* 2019;7:e1288–9.
83. Saar A, Läll K, Alver M, et al. Estimating the performance of three cardiovascular disease risk scores: the Estonian Biobank cohort study. *J Epidemiol Community Health* 2019;73:272–7.
84. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017;186:1026–34.

85. Casey JA, Schwartz BS, Stewart WF, et al. Using electronic health records for population health research: a review of methods and applications. *Annu Rev Public Health* 2016;37:61–81.
86. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ* 2013;246:e5595.
87. Denaxas SC, George J, Herrett E, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol* 2012;41:1625–38.
88. Quint JK, Herrett E, Bhaskaran K, et al. Effect of β blockers on mortality after myocardial infarction in adults with COPD: population based cohort study of UK electronic healthcare records. *BMJ* 2013;347:6650.
89. Dinesh A, Denaxas S, Nicholas O, et al. Neutrophil counts and initial presentation of 12 cardiovascular diseases: a CALIBER cohort study. *J Am Coll Cardiol* 2017;69:1160–9.
90. Uijl A, Koudstaal S, Direk K, et al. Risk factors for incident heart failure in age- and sex-specific strata: a population-based cohort using linked electronic health records. *Eur J Heart Fail* 2019;21:1197–206.
91. Leening MJG, Cook NR, Franco OH, et al. Comparison of cardiovascular risk factors for coronary heart disease and stroke type in women. *J Am Heart Assoc* 2018;7.
92. Pujades-Rodriguez M, George J, Shah AD, et al. Heterogeneous associations between smoking and a wide range of initial presentations of cardiovascular disease in 1937360 people in England: lifetime risks and implications for risk prediction. *Int J Epidemiol* 2015;44:129–41.
93. Herrett E, Shah AD, Boggon R. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. *BMJ* 2013;346:f2350.
94. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. *BMC Med Res Methodol* 2021;21:234.
95. Verheij RA, Curcin V, Delaney BC, et al. Possible sources of bias in primary care electronic health record data use and reuse. *J Med Internet Res* 2018;20:e185.
96. Williams BA, Voyce S, Sidney S, et al. Establishing a national cardiovascular disease surveillance system in the United States using electronic health record data: key strengths and limitations. *J Am Heart Assoc* 2022;11:24409.

97. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *The BMJ* 2009;338:157–60.
98. Kukhareva PV, Caverly TJ, Li H, et al. Inaccuracies in electronic health records smoking data and a potential approach to address resulting underestimation in determining lung cancer screening eligibility. *JAMIA* 2022;29:779–88.
99. Uusküla A, Oja M, Tamm S, et al. Prevaccination prevalence of type-specific human papillomavirus infection by grade of cervical cytology in Estonia. *JAMA Netw Open* 2023;6:e2254075.
100. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ* 2007;335:136.
101. Habicht T, Reinap M, Kasekamp K, et al. Eesti: Tervisesüsteemi ülevaade. *Tervisesüsteemid muutustes* 2023;25:i-204.
102. Oja M, Tamm S, Mooses K, et al. Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned. *JAMIA Open* 2023;6.
103. Choi JW, Lee J, Jung SJ, et al. Use of sedative-hypnotics and mortality: A population-based retrospective cohort study. *J Clin Sleep Med* 2018;14:1669.
104. Krittanawong C, Maitra NS, Qadeer YK, et al. Association of depression and cardiovascular disease. *Am J Med* 2023;136:881–95.
105. Go AS, Chertow GM, Fan D, et al. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* 2004;351:1296–305.
106. Telling K, Luman M, Rosenberg M, et al. Kroonilise neeruhaiguse ennetus ja käsitus (ajakohastatud). *Ravijuhend*. (<https://www.ravijuhend.ee/tervishoiuvarav/juhendid/250/kroonilise-neeruhaiguse-ennetus-ja-kasitus-ajakohastatud>). [04.05.2024].
107. SNOMED International. SNOMED CT. (<https://www.snomed.org/>). [12.02.2024].
108. Pencina MJ, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72.
109. OHDSI. PredictionComparison: an R package for comparing patient-level predictions and prediction models. (<https://github.com/OHDSI/PredictionComparison>). [17.04.2024].

110. Bots SH, Peters SAE, Woodward M. Sex differences in coronary heart disease and stroke mortality: a global assessment of the effect of ageing between 1980 and 2010. *BMJ Global Health* 2017;2:e000298.
111. Mikkola TS, Gissler M, Merikukka M, et al. Sex differences in age-related cardiovascular mortality. *PLoS ONE* 2013;8:e63347.
112. StackExchange. Calculate a 95% confidence interval and p-value for the change in C-statistic using bootstrap with R. (<https://stats.stackexchange.com/questions/104518/calculate-a-95-confidence-interval-and-p-value-for-the-change-in-c-statistic-us>). [02.05.2024].
113. Peters SAE, Woodward M, Rumley A, et al. Plasma and blood viscosity in the prediction of cardiovascular disease and mortality in the Scottish Heart Health Extended Cohort Study. *Eur J Prev Cardiol* 2017;24:161–7.
114. Li C, Liu X, Shen P, et al. Improving cardiovascular risk prediction through machine learning modeling of irregularly repeated electronic health records. *Eur Heart J Digital Health* 2024;5:30–40.
115. National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction, including lipid modification. NICE 2023. NICE Clinical Guidelines, No. 238.
116. Leening MJG, Vedder MM, Wittteman JCM, et al. Net reclassification improvement: computation, interpretation and controversies. *Ann Intern Med* 2014;160:122–31.
117. Van Calster B, Vickers AJ, Pencina MJ, et al. Evaluation of markers and risk prediction models: overview of relationships between NRI and decision-analytic measures. *Med Decis Making* 2013;33:490–501.
118. Binuya MAE, Engelhardt EG, Schats W, et al. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Med Res Methodol* 2022;316.
119. Vickers AJ, Calster BV, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016;352:i6.
120. Kallavus K, Laisaar KT, Rätsep A, et al. National lung cancer screening program feasibility study in Estonia. *Interact Cardiovasc Thorac Surg* 2023;36:i6.

Development of cardiovascular disease prediction models on Estonian national electronic health records, 2024

Laura Ilves

Summary

Background: Any personalized prevention service requires an accurate methodology to identify high-risk individuals. Having identified people at high risk, there is a need for sufficiently effective and accessible interventions, such as cholesterol-lowering drugs in the case of cardiovascular diseases (hereafter: CVD). Research to date in predicting CVD risk has focused on risk models that require a clinical visit and blood test. Comparatively little research has been done on how accurate risk assessment can be made using existing electronic health records. If such a model is precise enough, it would help future services to prioritize who to invite to further clinical trials. This, in turn, provides the basis for a cost-effective CVD screening program that has yet to be developed.

Objectives: To develop and internally validate two CVD risk prediction models to estimate the 5-year risk of primary cardiovascular disease in women and men.

Design: Retrospective cohort study.

Participants: A 10% random sample of the Estonian population ($n = 150\ 824$) from 2012 to 2019 and all their health data from three national health databases (claims, prescriptions, and electronic health records). 52 172 people aged 40–74 years were included in model development. Patients were free of cardiovascular disease at baseline.

Methods: The Cox proportional hazards model was used to create two models. The first model included age and gender. The second model used age, gender, type 1 diabetes, type 2 diabetes, hypertension, rheumatoid arthritis, atrial fibrillation, migraine, chronic kidney disease, severe mental illness, antidepressant use, corticosteroid use, sedative and hypnotics use. Continuous variables, such as systolic blood pressure, total cholesterol measurement, and body mass index, were also considered, but data missingness for these risk factors was $> 50\%$. Harrell's C statistic was determined in training and test sets separately. Both models were internally validated with 10-fold cross-validation. Net reclassification improvement was calculated to compare both models.

Results: For both models, 52 172 people were included with 1935 incident cases of cardiovascular disease. The first model's (age-gender model) C statistic for the train set was 0.750 (95% CI 0.738–0.761) and the test set was 0.747 (95% CI 0.727–0.768). The final model's C statistic for the train set was 0.762 (95% CI 0.751–0.773), and the test set was 0.757 (95% CI 0.737–0.778). Net reclassification improvement was +0.112 (95% CI 0.098–0.123) in favor of a new model with additional risk factors.

Conclusions: Two new CVD risk prediction models were developed and internally validated. Including non-continuous variables can help doctors identify those at most risk of heart disease and stroke and invite those people for additional analyses.

Tänuavaldus

Soovin siiralt tänada:

1. Juhendajaid Taavi Tillmann, Marek Oja ja Sirli Tamm – meeldiva koostöö, väärtuslike nõuannete ja võimaluse eest õppida parimatelt.
2. TÜ terviseinformaatika uurimisrühma – andmete kasutamise võimaldamise ja praktiliste kogemuste eest.
3. Kõiki lähedasi ja mõttekaaslasi – toetuse ja põnevate arutelude eest magistritöö teemadel.

Elulugu

Üldandmed

Ees- ja perekonnanimi: Laura Ilves
E-post: laura.ilves.99@gmail.com

Haridus

2022–2024 Tartu Ülikool, rahvatervishoid, magistriõpe
2019–2022 Tartu Ülikool, füsioteraapia, BSc
2015–2018 Tartu Miina Härma Gümnaasium

Keelteoskus

eesti keel emakeel
inglise keel C2
vene keel A2
prantsuse keel A2

Töökogemus

09.2023–04.2024 Tartu Ülikool, arvutiteaduse instituut, terviseinformaatika uurimisrühm, praktikant
05.2022–01.2023 Tartu Ülikooli Kliinikum, Ortopeediakliinik, füsioterapeut

Erialane enesetäiendus

2023 Rahvatervishoiu suvekool: „Rahvatervishoiu nüüdisaegsed probleemid ja väljakutsed“ (1,5 EAP)

Kuupäev: 27.05.2024

Lisad

Lisa 1. Töös kasutatavad SVH definitsioonid diagnooside kaupa

Tabel 1. Töös kasutatud esmase SVH definitsioon RHK-10 järgi

RHK-10 kood	Vastav nimetus
I21	äge müokardiinfarkt
I22	korduv müokardiinfarkt
I23	ägeda müokardiinfarkti mõned lähitüübid
I24	südame ägeda isheemiatõve muud vormid
I25	krooniline südame isheemiatõbi
I63	peajuinfarkt
I64	täpsustamata kas hemorraagia või infarktitekkene insult e rabandus

Tabel 2. Töös kasutatud eelneva SVH ehk kohordist välja arvamise SVH definitsioon RHK-10 järgi

RHK-10 kood	Vastav nimetus
G45	peaaju transitoorse isheemia atakid ja sellesarnased sündroomid
I20	stenokardia e rinnaangiin e rinnaahistus
I21	äge müokardiinfarkt
I22	korduv müokardiinfarkt
I23	ägeda müokardiinfarkti mõned lähitüsistused
I24	südame ägeda isheemiatõve muud vormid
I25	krooniline südame isheemiatõbi
I60	subarahnoidaalne hemorraagia e ämblikuvõrkkesta-alune verevalum
I61	intratserebraalne hemorraagia e peaajusisene verevalum
I62	muu mittetraumaatiline koljusisene verevalum
I63	peaajuinfarkt
I64	täpsustamata kas hemorraagia või infarktitekkene insult e rabandus
I65	pretserebraalarterite peaajuinfarktita oklusioon e sulgus ja stenoos e ahenemus
I66	peaajuarterite peaajuinfarktita oklusioon ja stenoos
I67.2	peaaju ateroskleroos, peaaju ateroomid
I69	peaajuveresoonte haiguse jääknähud
I70	ateroskleroos
I73.9	perifeersete veresoonte täpsustamata haigused

Lisa 2. Mudeli tunnuste definitsioonid

Tabel 3. Mudelisse kaasatud tunnused koos tunnuse tüübi ja definitsiooniga

Tunnus	Tunnuse tüüp	Definitsioon
Demograafilised tunnused		
Sugu	Binaarne	Mees/naine
Vanus	Pidev	40–74aastased inimesed
Diagnoosid		
Hüpertensioon	Binaarne	I10–15
Kodade virvendus ja laperdus	Binaarne	I48
Krooniline neeruhaigus	Binaarne	N18.3–18.5 või ≥ 2 mõõtmistulemust, kus eGFR < 60 mL/min/1,73m ² vähemalt 3kuulise vahega
Migreen	Binaarne	G43
Raske vaimne haigus	Binaarne	F20–31, F32.3, F33.3
Reumatoidartriit	Binaarne	M05–06
1. tüüpi diabeet	Binaarne	E10
2. tüüpi diabeet	Binaarne	E11
Ravimite tarvitamine		
Antidepressandid	Binaarne	ATC-puu alusel N06A rühma kuuluvad toimeained
Kortikosteroidid	Binaarne	ATC-puu alusel C05AA, R01AD ja S03BA rühma kuuluvad toimeained
Uinutid ja rahustid	Binaarne	ATC-puu alusel N05C rühma kuuluvad toimeained

Lisa 3. Mudelite koefitsiendid

Mudelite koefitsiendid, mille põhjal arvutati riskimäärade suhted, on välja toodud Tabelis 4.

Tabel 4. Vanus-soo mudeli ja lõpliku mudeli koefitsiendid koos usaldusvahemikega

Tunnus	Mudel 1	Mudel 2
	Vanus-soo mudel	Lõplik mudel
	Koefitsient (95% CI)	Koefitsient (95% CI)
Demograafilised tunnused		
Mees	0,803 (0,698; 0,909)	0,832 (0,725; 0,940)
Naine	0,000*	0,000
Vanus	0,090 (0,085; 0,096)	0,079 (0,073; 0,085)
Diagnoosid		
Hüpertensioon		0,434 (0,318; 0,550)
Kodade virvendus ja laperdus		0,391 (0,197; 0,575)
Krooniline neeruhaigus		0,055 (-0,229; 0,752)
Migreen		0,000**
Raske vaimne haigus		0,012 (-0,199; 0,410)
Reumatoidartriit		0,343 (0,020; 0,638)
1. tüüpi diabeet		0,104 (-0,189; 0,496)
2. tüüpi diabeet		0,344 (0,188; 0,496)
Ravimite tarvitamine		
Antidepressandid		0,083 (-0,102; 0,306)
Kortikosteroidid		0,000**
Uinutid ja rahustid		0,083 (-0,054; 0,233)

*Võrdlusrühm

**Lasso meetodiga mudelist eemaldatud

Lisa 4. Riskikalkulatsioon kahe stsenaariumi näitel lõplikku mudelit kasutades

Stsenaarium 1. 58aastane mees, kellel on kaasvalt hüpertensiooni ning kodade virvenduse diagnoos, tarvitab rahusteid (Tabel 5).

Tabel 5. Näidis riskikalkulatsioon lõpliku mudeliga, stsenaarium 1

Riskitegur	HR	Koefitsient	Stsenaarium	Selle isiku	Riskikomponent
				väärtused	
Sugu	2,30	0,832	Mees	1	0,832
Vanus	1,08	0,079	58	18	1,422
Hüpertensioon	1,54	0,434	Jah	1	0,434
Kodade virvendus	1,49	0,391	Jah	1	0,391
Krooniline neeruhaigus	1,06	0,055	Ei	0	0,000
Raske vaimne haigus	1,01	0,012	Ei	0	0,000
Reumatoidartriit	1,41	0,345	Ei	0	0,000
1. tüüpi diabeet	1,11	0,104	Ei	0	0,000
2. tüüpi diabeet	1,41	0,344	Ei	0	0,000
Antidepressandid	1,09	0,083	Ei	0	0,000
Uinutid ja rahustid	1,09	0,083	Jah	1	0,083

Mudeli baaselulemus = 0,9626

Mudeli baasrisk = 1 – baaselulemus = 1 – 0,9626 = 0,0374 ehk 3,74%

Riskikomponentide summa = soo komponent + vanuse komponent + ... = 3,162

Eksponeeritud riskikomponentide summa = $e^{3,162} = 23,618$

Elulemuse tõenäosus = baaselulemus^{eksponeeritud riskikomponentide summa} = $0,9626^{23,618} = 0,406$ ehk 40,6%

SVH tekke tõenäosus 5 aasta jooksul = 1 – elulemuse tõenäosus = 1 – 0,406 = 0,594 ehk **59,4%**

Stsenaarium 2. 42aastane naine, kellel on kaasvalt raske vaimse haiguse diagnoos ning tarvitab antidepressante (Tabel 6).

Tabel 6. Näidis riskikalkulatsioon lõpliku mudeliga, stsenaarium 2

Riskitegur	HR	Koefitsient	Stsenaarium	Selle isiku väärtused	Riskikomponent
Sugu	2,30	0,832	Naine	0	0,000
Vanus	1,08	0,079	42	2	0,158
Hüpertensioon	1,54	0,434	Ei	0	0,000
Kodade virvendus	1,49	0,391	Ei	0	0,000
Krooniline neeruhaigus	1,06	0,055	Ei	0	0,000
Raske vaimne haigus	1,01	0,012	Jah	1	0,012
Reumatoidartriit	1,41	0,345	Ei	0	0,000
1. tüüpi diabeet	1,11	0,104	Ei	0	0,000
2. tüüpi diabeet	1,41	0,344	Ei	0	0,000
Antidepressandid	1,09	0,083	Jah	1	0,083
Uinutid ja rahustid	1,09	0,083	Ei	0	0,000

Mudeli baaselulemus = 0,9626

Mudeli baasrisk = 1 – baaselulemus = 1 – 0,9626 = 0,0374 ehk 3,74%

Riskikomponentide summa = tegur 1 + tegur 2 + ... = 0,253

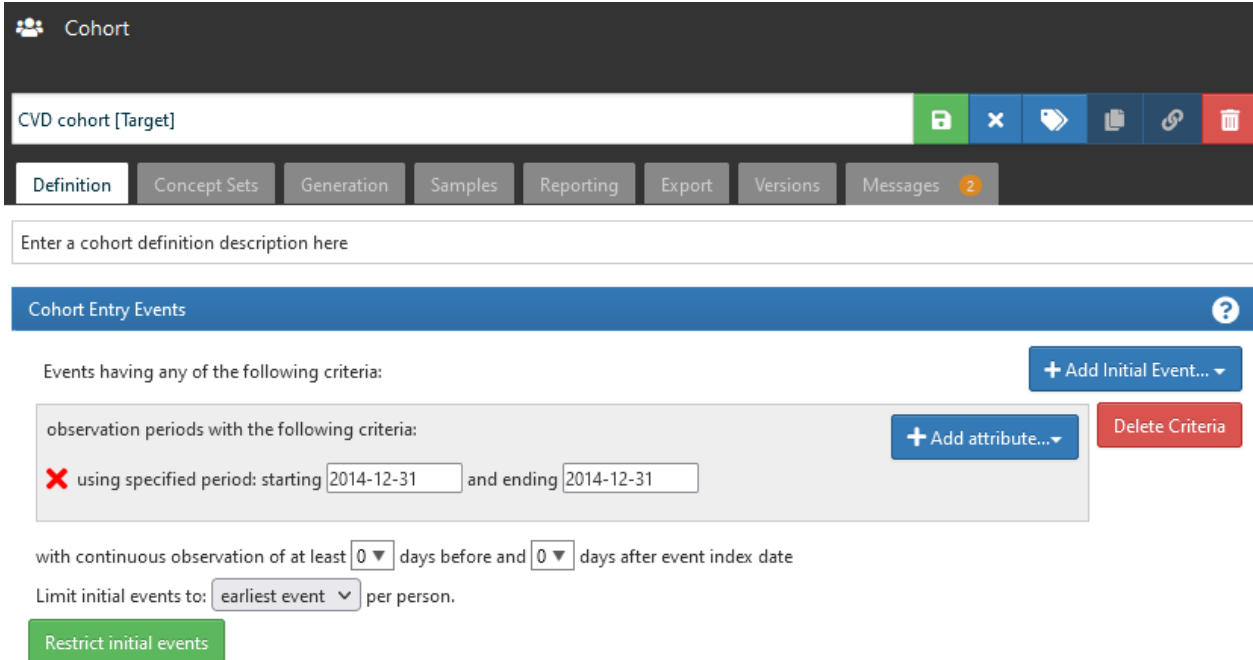
Eksponeeritud riskikomponentide summa = $e^{0,253} = 1,288$

Elulemuse tõenäosus = baaselulemus^{eksponeeritud riskikomponentide summa} = $0,9626^{1,288} = 0,952$ ehk 95,2%

SVH tekke tõenäosus 5 aasta jooksul = 1 – elulemuse tõenäosus = 1 – 0,952 = 0,048 ehk **4,8%**

Lisa 5. Kohordi defineerimine tarkvaraga ATLAS

Järgnevalt on välja toodud näidisenä 40–74aastaste eelneva SVHta inimeste kohordi defineerimine ATLASes. Esimese etapina tuleb täita kohorti sisenemise tingimus (*Cohort Entry Events*). Joonisel 1 defineeritakse kohort, kuhu kuuluvad kõik patsiendid, kes on andmebaasis olemas seisuga 31.12.2014.



The screenshot shows the ATLAS Cohort Entry Events configuration screen. At the top, there is a header 'Cohort' and a search bar containing 'CVD cohort [Target]'. Below the search bar is a navigation menu with tabs: Definition, Concept Sets, Generation, Samples, Reporting, Export, Versions, and Messages (with a notification badge). A text input field for 'Enter a cohort definition description here' is present. The main section is titled 'Cohort Entry Events' and contains the following configuration options:

- Events having any of the following criteria: [+ Add Initial Event...](#)
- observation periods with the following criteria: [+ Add attribute...](#) [Delete Criteria](#)
- using specified period: starting and ending
- with continuous observation of at least days before and days after event index date
- Limit initial events to: per person.
- [Restrict initial events](#)

Joonis 1. Kohorti sisenemise tingimuste defineerimine.

Teiseks tuleb lisada kohorti kaasamise kriteeriumid (*Inclusion Criteria*). Antud magistritöö raames kaasati inimesed, kes on jälgimisaja alguses ehk seisuga 31.12.2014 40–74-aastased (Joonis 2) ning kellel puudus eelnev SVH diagnoos (Joonis 3). Seega kohorti kaasati 40–74aastased inimesed (*Age 40–74*), kellel pole esinenud varasemat SVH diagnoosi (*No prior CVD diagnosis*).

Inclusion Criteria ?

New inclusion criteria Age 40-74 Copy Delete

1. Age 40-74

2. No prior CVD diagnosis having all of the following criteria: + Add criteria to group... ▾

with the following event criteria:

+ Add attribute... ▾

✗ with age Between 40 and 74 Delete Criteria

Limit qualifying events to: all events per person.

Joonis 2. Kohorti kaasamise kriteeriumite defineerimine, 1. osa.

Inclusion Criteria ?

New inclusion criteria No prior CVD diagnosis Copy Delete

1. Age 40-74

2. No prior CVD diagnosis having all of the following criteria: + Add criteria to group... ▾

with exactly 0 using all occurrences of:

a condition occurrence of Exclusion CVDs + Add attribute... ▾

where event starts between

All days Before and 0 days After index start date add

additional constraint

The index date refers to the event from the Cohort Entry criteria.

restrict to the same visit occurrence

allow events from outside observation period

Delete Criteria

Limit qualifying events to: all events per person.

Joonis 3. Kohorti kaasamise kriteeriumite defineerimine, 2. osa.

Lõpuks tuleb määrata kohordist väljumise kriteeriumid (*Cohort Exit*). Antud juhul oli selleks jälgimisaja lõpp (*Event Persistence: end of continuous observation*) ehk 31.12.2019.

Cohort Exit ?

Event Persistence:
Event will persist until: end of continuous observation

Censoring Events:
Exit Cohort based on the following criteria: + Add Censoring Event... ▾

No censoring events selected.

Joonis 4. Kohordist väljumise kriteeriumite defineerimine.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Laura Ilves,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Südame-veresoonkonna haiguste riskimudelite loomine Eesti elektroonilistel terviseandmetel“, mille juhendajad on Taavi Tillmann, Marek Oja ja Sirli Tamm, 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.

Laura Ilves

27.05.2024