TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT
GENOOMIKA INSTITUUT
INIMESE HAIGUSTE GENOOMIKA TÖÖGRUPP

Eesnäärmevähi riskimudeli arendamine Eesti terviseandmete põhjal

Magistritöö
30 EAP

Telver Objärtel

Juhendajad: PhD Mart Kals
MD Anu Reigo
Kaasprofessor, PhD Raivo Kolde

TARTU 2024

Infoleht

Eesnäärmevähi riskimudeli arendamine Eesti terviseandmete põhjal

Eesnäärmevähk on meeste seas üks kõige levinum ja üks kõige suurema suremusega kasvaja, mida on varajastes staadiumites raske tuvastada. Selles väitekirjas koostati statistiline mudel, mis ennustab 40–59-aastaste meeste eesnäärmevähiriski. Mudeli loomisel kasutati Eesti meeste terviseandmeid perioodist 2012–2019 (RITA MAITT andmebaas). Mudelit valideeriti Eesti geenivaramu kohordis (30 045 meest, kellest 622 olid juhud). Parim mudel kasutab riski ennustamisel teavet vanuse, eesnäärme suurenemuse ja kroonilise prostatiidi kohta. Valideerimisandmestikul oli mudeli 5–aastane ennustusvõime hea ($AUROC = 0{,}852$; 95% usaldusintervall $84{,}1-86{,}2$; $AUPRC = 7{,}83$). Eesnäärmevähi polügeenne riskiskoor ei suurendanud mudeli ennustusvõimet. Mudelit saab kasutada PSA mõõtmise või eesnäärme muu kliinilise uuringu näidustamiseks.

Märksõnad: eesnäärmevähk, riskimudel, varajane avastamine, riski hindamine, sõeluuring

CERCS: B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika
B220 Geneetika, tsütogeneetika
B680 Rahvatervishoid, epidemioloogia

Development of a prostate cancer risk model on Estonian health data

Prostate cancer (PCa) is one of the most frequent cancer types in men. It is notably difficult to detect in its early stages. We developed a PCa risk model for 40-59-year-old men using Estonian health data of 16 506 men collected from 2012 to 2017 in the RITA MAITT database. The model was validated in 30 045 unaffected men (out of which 622 were later diagnosed with PCa) recruited from 2003 to 2022 to the independent Estonian Genome Centre study. The best model's predictors are benign prostatic hyperplasia and chronic prostatitis. The model discriminated well ($AUROC = 0{,}852$; 95%CI 84,1 to 86,2; $AUPRC = 7{,}83$). A polygenic risk score for PCa did not help to discriminate incident PCa better than the original model. After pilot testing and evaluation in a clinical trial, the model can be used to indicate PSA testing or prostate examinations in clinical settings.

Keywords: prostate cancer, risk model, early detection, risk stratification, screening

CERCS: B110 Bioinformatics, medical informatics, biomathematics, biometrics
B220 Genetics, cytogenetics
B680 Public health, epidemiology

Sisukord

Kasutatud lühendid

AUROC	ROC-kõvera alune pindala (*area under ROC curve*)
AUPRC	täpsus-saagis kõvera alune pindala (*area under precision-recall curve*)
CDM	ühine andmemudel (*common data model*)
EGV	Tartu Ülikooli Eesti geenivaramu
ES	eesnäärme healoomuline suurenemine (*benign prostatic hyperplasia*)
ENV	eesnäärmevähk (*prostate cancer*)
GWAS	geneetiline assotsiatsiooniuuring (*genome-wide association study*)
LD	aheldustasakaalutus (*linkage disequilibrium*)
OHDSI	*Observational Health Data Sciences and Informatics*
OMOP	*Observational Medical Outcomes Partnership*
OR	šansside suhe (*odds ratio*)
PLP	patsienditaseme ennustus (*patient-level prediction*)
PRS	polügeenne riskiskoor (*polygenic risk score*)
PSA	prostataspetsiifiline antigeen (*prostate-specific antigen*)
RHK-10	rahvusvaheline haiguste klassifikatsioon, 10. versioon (ICD-10)
SNOMED	*Systematized Medical Nomenclature for Medicine*
SNP	ühenukleotiidne polümorfism (*single nucleotide polymorphism*)
WHO	Maailma Terviseorganisatsioon (*World Health Organization*)

Sissejuhatus

Inimesed pole kunagi nii kaua elanud kui praegu, 21. sajandi algul. Arenenud maailm on saavutanud viimase saja aastaga ligi kahekordse inimese elueapikkuse tõusu. Probleem on nüüd eluea pikkuse asemel pigem selles, et vähk kipub seda pikemat eluiga märkimisväärselt rohkem mõjutama. Inimajaloo kõige pikemate elude jooksul võiks olla rohkem tervelt elatud aastaid. Üks haigus, mis küll üleöö ei tapa, kuid teeb põdeja elu igal juhul oluliselt ja progresseruvalt halvemaks (ja lühemaks), on vähk. Nii nagu enamikus maailmas, on ka Eesti meeste seas eesnäärmevähk kõige sagedasem pahaloomuline kasvaja – ligi 30% kõikidest meessoo uutest vähijuhtudest. Paljud küll elavad veel peale esmadiagnoosi edasi, kuid sageli jääb neid jääb elu lõpuni saatma tohutu vaimne ja füüsiline vaev, mis kaasneb vähi ja selle ravi tagajärgedega. Eesnäärmevähi puhul väärib märkimist veel seni pigem ebaintuitiivne asjaolu, et eriti vanematel meestel kiputakse seda üle diagnoosima ja üle ravima, samas kui noorematel meestel jääb kliiniliselt oluline vähk sageli tuvastamata, kuni on tervenemise väljavaateks juba liiga hilja. Ühtlasi on näha, et vähk ei ole enam ainult vanema ea haigus, nagu kaua on arvatud. Üha enam tuvastatakse vähki ka keskealistel ja isegi noortel.

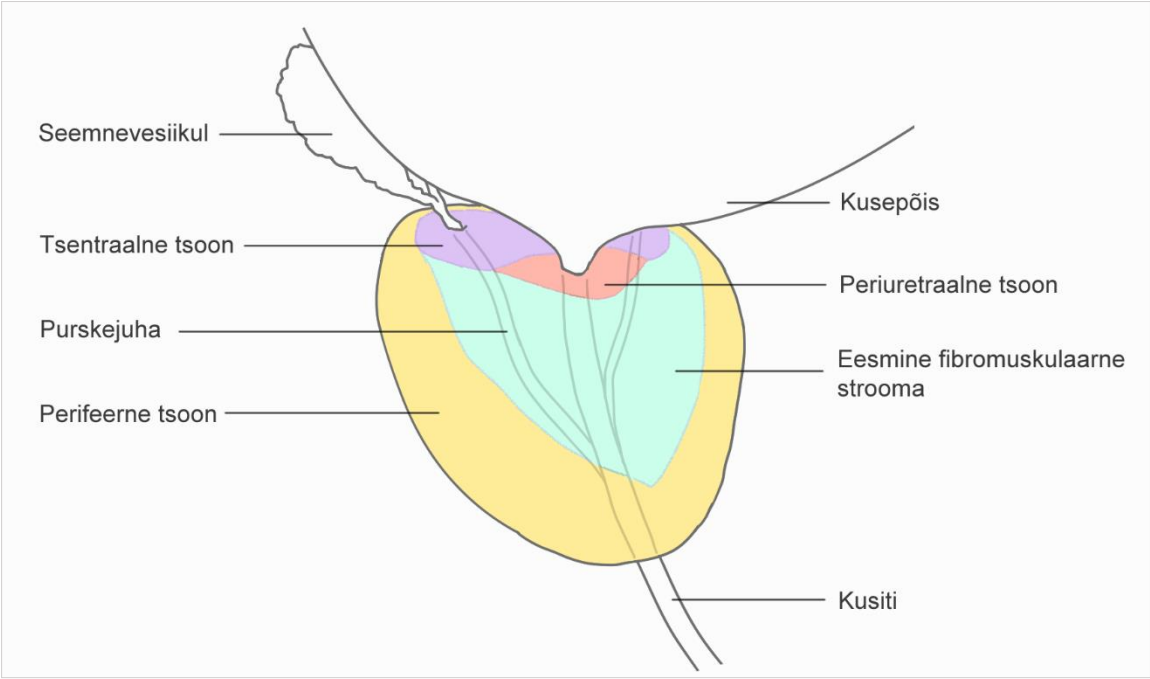
Eesnäärmevähi varajane avastamine on lahtine probleem. Selle väitekirja eesmärk on saada teada, kas kliiniline vaatlusteave suudab ennustada uute eesnäärmevähijuhtude teket Eesti meeste seas. Kliinilisel vaatlusteabel põhineva mudeli potentsiaalne kasu aitaks suurema vähiriskiga mehi varem arstlikku jälgimisse suunata.

Uurimustöö koostati Tartu Ülikooli genoomika instituudis (inimese haiguste genoomika töögrupis) ja Tartu Ülikooli arvutiteaduse instituudis (terviseinformaatika uurimisrühmas). See väitekiri on osa rahvusvahelisest teadusprojektist “OPTIMA – Optimaalne ravi kasvajatega patsientidele Euroopas tehisintellekti abil” (01.10.2021–30.09.2026). Eesti terviseandmete statistiliseks analüüsiks kasutati Tartu Ülikooli teadusarvutuste keskust, UT HPC.

1. KIRJANDUSE ÜLEVAADE

1.1. Eesnääre

Eesnääre (ingl *prostate*) on meessoo organ. See on umbes kastanimuna suurune ja asub vahetult mehe kusepõie all, ümber kusiti (joonis 1). Selle üks funktsioon on toota vedelikku, mis seguneb munandites toodetud seemnerakkudega ja seeläbi aitab moodustada seemnevedelikku. Lisaks on eesnäärmel ka immuunroll. Eesnäärme mõlemas otsas on sulgurlihased, mis aitavad reguleerida kusemist ja seemnepurset (IQWiG, 2022).



Joonis 1. Inimese eesnääre ja selle koetsoonid, vaadatuna eespoolt. Eesnäärme südamikus ümbritseb kusitit ka üleminekutsoon. Munandites toodetud seemnerakud liiguvad seemnepurske jaoks mööda seemnejuhasid eesnäärmesse, kus need segunevad eesnäärmevedelikuga. Sealt liigub äsjatekkinud sperma edasi peenisesse. Originaaljoonis.

Kolm kõige sagedasemat eesnäärmehaigust on prostatiit (tuntud kui ka eesnäärmepõletik, kuigi mõne prostatiidi vormi puhul ei ole eesnääre ise põletikus), eesnäärme healoomuline suurenemine (edaspidi EHS) ning eesnäärmevähk. Umbes viiendik eesti 20–50 a vanustest meestest Eestis põevad eesnäärmepõletikku ning >50% üle 50-aastastest meestest kurdavad peamiselt EHS-st tingitud kusemishäirete üle. Peamine ühine nimetaja on alumiste kuseteede sümptomid (*lower urinary tract symptoms*). Need jagunevad omakorda kaheks (PUNAB, 2023):

1. Ärritusega seotud kusemishäired – ebamugavustunne kusemisel, sagenenud kusemine, pakiline kusemisvajadus, öine kuselkäimine, ebamugavustunne seemnepurske ajal või järel. Need viitavad pigem põletikulistele haigustele.
2. Takistustüüpi kusemishäired – nõrk kusejuga, takistus kusemise alustamisel, katkendlik kusevool, põie mittetäielik tühjenemine, vajaduse kusemise alguses pingutada. Need on iseloomulikud just eesnäärme suurenemisele.

Samuti viitavad eesnäärmeprobleemidele erektsioonihäired ning veri uriinis või seemnevedelikus. Ka väga levinud haigused nagu südamehaigused, närvihaigused ja diabeet võivad soodustada teatud tüüpi kusemishäirete teket. Ühtlasi võib eesnäärmeprobleemide riski suurendada seksuaalsuhete arv, eriti kui need on kaitseta. (PUNAB, 2023)

1.2. Eesnäärmevähk

Ühel mehel seitsmest diagnoositakse elu jooksul eesnäärmevähk. 112 riigi (sealhulgas Eesti) meeste seas on eesnäärmevähk kõige sagedasem pahaloomuline kasvaja. Vähihaigete meeste seas on eesnäärmevähk teine kõige suurema suremusega vähk, esikohal on kopsuvähk. Igal aastal moodustab eesnäärmevähk ligi kolmandiku kõikidest Eesti meeste vähi esmajuhtudest. (JAMES jt, 2024; TERVISE ARENGU INSTITUUT, 2023; EUROOPA VÄHITEABESÜSTEEM, 2023).

Eesnäärmevähk (*prostate cancer*, lühidalt **ENV**) on komplekshaigus – selle avaldumist mõjutavad korraga mitmed geneetilised ja mittegeneetilised tegurid. Suurimad riskitegurid on:

- vanus – mida vanem mees, seda suurem risk;
- rahvus (*ethnic origin*) – Ameerika mustanahalistel meestel on 2,2 korda suurem tõenäosus surra eesnäärmevähki kui valgenahalistel (ACS, 2019);
- geneetiline eelsoodumus, mida sageli hinnatakse kaudselt pere haigusloo kaudu (*family history*), nt kas vähihaige sugulastel on olnud ENV, sugulaste vanus esmadiagnoosi saamisel jne. Kui isal või vennal on diagnoositud eesnäärmevähk, on risk vähki haigestuda kuni 2 korda suurem kui mehel, kelle perekonnas haigust ei esine.

Kui diagnoosihetkel vähk ei ole metastaseerunud, siis ühe aasta suhteline elulemus (*survival*) on 98%, viie aasta suhteline elulemus 94% ning kümne aasta suhteline elulemus on 91%. Samas: kui diagnoosihetkel on vähk juba metastaseerunud, siis ühe aasta elulemus on 82%, viie

aasta elulemus on 34% (tõenäosus viie aasta jooksul surra on 2/3) ja kümne aasta elulemus on 20%. (TAI, 2023).

1.2.1. Sümptomid

Eesnäärmevähk on üldiselt aeglase kuluga ning selle varastes arengustaadiumites enamasti asümptomaatiline. Haigus võib algfaasis ilma selgete tunnusteta areneda aastaid. Lokaalne progressioon võib põhjustada: (PUNAB, 2006; CANCER RESEARCH UK, 2022)

- alumiste kuseteede sümptomeid (*lower urinary tract symptoms*);
- erektsioonihäireid;
- vaagnapiirkonna valu;
- vere esinemist kuses või seemnevedelikus;
- luuvalu, seljavalu ja/või reievalu, kui eesnäärmevähk on metastaseerunud (kaugelearenenud).

Seega kõigi levinud eesnäärmehaiguste, sealhulgas eesnäärmevähi puhul võivad tekkida kohati sarnased ja ebaiseloomulikud sümptomid, kui üldse. Eesnäärmevähi tuvastamiseks on vaja kasutada täiendavaid meetmeid.

1.2.2. Diagnostika

Kliinilises praktikas kasutatavad diagnostilised meetodid (tabel 1, järgmine lk).

Tabel 1. Euroopas, sh Eestis kasutatavad eesnäärmevähi analüütikameetodid (VESKIMÄGI jt 2020; EUROPEAN ASSOCIATION OF UROLOGY, 2024).

Meetod	Kirjeldus
Digitaalne rektaalne palpatsioon (*digital rectal examination* ehk DRE)	Meetod, kus meditsiinitöötaja katsub sõrmega päraku kaudu eesnääret. Umbes ~18% juhtudel tuvastatakse ainuüksi sellega ENV. Meetodi tundlikkus ja ennustusväärtus eesnäärmevähi varajasel avastamisel on väike. Kasutatakse edasiste uuringute (nt MRT või biopsia) näidustamiseks.
Transrektaalne ultraheliuuring (*transrectal ultrasound* ehk TRUS)	Laialt kasutatav, kuid ebatäpne meetod. Ei ole piisavalt usaldusväärne, et iseseisvalt diagnostikaks kasutada.
Prostataspetsiifiline antigeen ehk PSA	Vereproovist määratava PSA mõõtetulemuse abil määratakse riskikategooria. Suure PSA mõõtetulemuse puhul otsustatakse ka, kas tegu võib olla lokaalse vähiga või lokaalselt levinud vähiga (*locally advanced cancer*). Sellest lähemalt järgmises alampeatükis.
Pere haiguslugu (*family history*, lühend FH)	ENV risk suureneb, kui perekonnas on varem esinenud eesnäärmevähki või rinnavähki. Pere varasemate haigete arvessevõtmine on kaudne meetod (*proxy*), et hinnata geneetilist eelsoodumust. Kuigi see on veel praegune standard, siis uuemad uuringud soovitavad konkreetselt uuritava isiku genotüpiseerimist.
Magnetresonantstomo-graafia ehk MRT (*magnetic resonance imaging*)	Peamiselt kasutatakse edasise meetmena PSA- ja/või DRE-põhise ENV kahtluse korral. Aitab (liigseid) biopsiaid vältida. Lootustandev ja aktiivselt uuritav meetod. MRT põhjal näidustatud biopsiad on suurendanud diagnostilist täpsust. Kasutatakse ka olemasoleva vähi klassifitseerimisel.
Biopsia	Biopsia on eesnäärme koeproovi histopatoloogiline analüüs. Biopsiaga vaadatakse, kui sarnane on vähikahtlusega mehe eesnäärme histoloogiline preparaat terve eesnäärmega mehe koepildile. Biopsia põhjal määratakse patsiendile Gleasoni skoor. ENV-d üldiselt ilma biopsiata ei diagnoosita.

Euroopa Uroloogiaassotsiatsioon (*European Association of Urology*) on koostanud eesnäärmevähi diagnoosi- ja ravijuhendi (*Classification and Staging Systems*). Kõige uuemas versioonis (aprill 2024) on püstitatud ülaltoodud analüüsimeetodite põhjal diagnoosirada (*diagnostic pathway*), kus nii sümptomaatilistele kui ka asümptomaatilistele meestele soovitatakse koostada algne riskihinnang (*initial risk assesment*) PSA, DRE, pere haigusloo ja rahvuse põhjal.

- Kui risk on madal, siis planeeritakse järgmisi samme sõltuvalt olukorrast (*individualized follow-up*).
- Kui risk on keskmine, siis tuleb kasutada täiendavat riskilahutust, näiteks mõni riskikalkulaator, MRT ja/või kuse- või vereproov. Kui siit leitakse, et risk on madal, jäetakse mees jälgimisele. Kui risk on ainuüksi riskiarvutuse põhjal suur, siis tehakse MRT (kui MRT-d juba enne ei tehtud). Kui MRT tulemus on positiivne, siis tuleb mehele teha biopsia.

- Kui risk on algse riskihinnangu põhjal kõrge (PSA >50 ng/ml), tuleb mehele teha biopsia

1.2.2.1. PSA

Prostataspetsiifiline antigeen ehk PSA on ensüüm, mida sekreteerivad eesnäärme epiteelirakud. Selle peamine roll on seemnevedelikus luua sobiv keskkond seemnerakkudele, kuid seda leidub ka vähesel määral vereseerumis. Kuna see on eesnäärme-spetsiifiline biomarker, siis sellel on eesnäärmeprobleemide riskilahutuse potentsiaal (MARTIN jt, 2024).

PSA mõõtmine on olnud ajalooliselt väga vastuoluline meetod, kuna see ei ole vähispetsiifiline, vaid kajastab ka teisi eesnäärmemuutusi, mis on tingitud näiteks prostatiidist, eesnäärme healoomulisest suurenemisest ja seksuaalvahekordadest (MCNALLY jt, 2020; MÜLLER jt, 2022). Kuigi PSA-põhine skriinimine on vähendanud eesnäärmevähist tingitud suremust (*prostate cancer specific mortality*), on see samal ajal suurendanud kliiniliselt ebaolulise eesnäärmevähi ülediagnoosimist ja üleravimist (ILIC jt, 2018; VOS jt, 2023). See omakorda tähendab ebavajalikke biopsiaid ja tervishoiu ressursside kulutamist ebaohtlikele eesnäärmevähkidele. Paljudes arenenud riikides, sealhulgas Eestis, ei ole seepärast riiklikku ENV-sõeluuringuprogrammi, see-eest soovi korral mehed saavad siiski end PSA suhtes testida. **Kõrge sissetulekuga riikides testitakse praegu PSA taset ebasüsteemselt ja asümmeetriliselt, mis päädib võimaliku kliinilise kahjutegemise ja vähese kliinilise kasuga, seejuures testitakse liialt vähe neid, kes vajaksid seda rohkem (40-70 aastased)** (VICKERS jt, 2023; PUNAB, 2023b; TASA jt, 2020).

1.2.2.2. Genotüüp

Eesnäärmevähi päritavus (*heritability*) on umbes 60%, samas kui muude vähitüüpide keskmine päritavus on ligi 33%, ehk eesnäärmevähi kujunemisel mängib isiku geneetiline profiil ehk genotüüp võrreldes teiste vähkidega erakordselt suurt rolli (NCI, 2024; REBBECK, 2016; HJELMBORG jt, 2014). Suures pildis saab geneetilised riskilookused jaotada nelja gruppi, kus esimesed kaks gruppi on seotud DNA reparatsiooniga (VIETRI jt, 2021; SANDHU jt, 2021; SEIBERT jt, 2023).

1. **DSB-geenid** (*double strand breaks*). Geenid, mis on seotud DNA kaksikahelaliste katkemiste parandamisega. Suurem osa nendest geenidest on seotud homoloogilise

rekombinatsiooniga: *BRCA2, BRCA1, ATM, PALB2, CHEK2, BRP1, NBS1*. Üks geen, *SPOP*, on seotud ka mittehomoloogsete DNA-otste ühendamisega.

2. **MMR-geenid** (*mismatch repair*). Geenid, mis reguleerivad DNA valepaardumiste parandamist:. *MLH1*, *MSH2*, *MSH6* ja *PMS2*.
3. **Kasvajate supressorgeenid** (*tumor suppressor gene*). *RB1*, *TP53*, *PTEN*.
4. **Muud geenid**. *HOXB13, PCA3, FOXA1, AR.*

Vähi ja muude komplekshaiguste geneetilised riskitegurid jagunevad kaheks: somaatilised mutatsioonid *vs.* iduteemutatsioonid (*germline mutations*). Somaatilised mutatsioonid tekivad peale sügoodi teket inimese eluea jooksul üle kogu keha väljaspool sugurakke, ja need seega ei pärandu hiljem järglastele. Need mõjutavad ainult kudesid, kus somaatiline mutatsioon on tekkinud. Iduteemutatsioonid on muutused sugurakkude DNA-s, mis saavad päranduda järglastele ja mis seega saavad mõjutada tervet organismi. (NHS, 2022) Kaasaegses kliinilises praktikas paraku vähiriski hindamiseks rutiinset geenitestimist veel ei kasutata, tuginetakse pigem kvalitatiivsele küsitlusele varasemate vähihaigete kohta uuringualuse perekonnas ja/või suguvõsas. See on kaudne informatsioon pärilikest idutee riskivariantidest, kuid ei ole kindel teadmine riskivariantide olemasolu kohta sellel indiviidil.

1.2.3. Varajane avastamine

Praegune vabatahtlik PSA-põhine süsteem ei ole piisavalt tõhus, et vähki varakult avastada. Eesnäärmevähi populatsioonipõhise tuvastamise puhul on suur probleem ülediagnoosimine. Kuigi eesnäärmevähk ei pruugi paljudel juhtudel olla eluohtlik, siis paraku leiab aset kliiniliselt ebaoluliste (asümptomaatiliste ja/või mitteagressiivsete) juhtude ülediagnoosimine, mis päädib üleravimisega. Diagnostika ja isegi diagnoosi teadmine põhjustab füüsilist ja vaimset traumat, eriti kui vähk on mitteagressiivne ja on ebatõenäoline, et see vähendaks mehe elukvaliteeti (seda näiteks eriti vanematel meestel). Seevastu potentsiaalselt surmava juhu korral saadakse diagnoos liiga hilja, mis on paraku üha sagenev liigvarajase ENV-suremuse põhjustaja (JAMES jt, 2024). Parim meetod vähisurmade vähendamiseks on vähi tekke vältimine, kuid selleks toimivaid soovitusi anda on seni keeruline. Paremuselt teine meede, mida peab koos vältimisega rakendama, on vähi või vähieelsete seisundite varajane avastamine. (PADRIK, 2020)

Vähi tuvastamise efektiivsuse suurendamiseks, üle- ja alaravi vähendamiseks ning ressursside mõistlikuks kasutamiseks tuleks luua riiklik riskikohandatud sõeluuringuprogramm, kus **sõeluuringu sihtrühma kitsendatakse personaalse riski põhjal** (VESKIMÄE jt, 2020; PADRIK, 2020; VICKERS jt, 2023; EAU, 2024).

1.3. Riski hindamine

Mis on risk? Üks definitsioon on, et risk on määramatus, mille tulemusel võib tekkida kahju. Sellest lähtuvalt: haigusrisk on võimalus, et inimene haigestub kindla vaatlusperioodi jooksul. Riski üldiselt väljendatakse tõenäosusena, mis jääb alati 0 ja 1 vahele. Et ennustada midagi tegeliku elu kohta, näiteks haigusriski, kasutatakse teaduses statistilist mudeldamist. **Statistiline mudel** on tavaliselt matemaatiline funktsioon või funktsioonide komplekt, mis sõltumatute tunnuste X põhjal ennustab uuritavat tunnust Y. Näiteks: sõltumatu tunnus on kehakaal ja uuritav tunnus on haigusrisk. Selline mudel võtaks sisendiks inimese kehakaalu ja tagastaks kehakaalule vastava haigusriski. **Riskimudel** on statistiline mudel, kus uuritav tunnus Y on haigusrisk, teisisõnu mudeliga ennustatakse tõenäosust, et inimene jääb mingil kindlal vaatlusperioodil haigeks.

Personaalse riski hindamiseks on loodud mitmeid riskimudeleid. On olemas vähemalt 100 ENV riskimudelit (ALADWANI jt, 2020). Väga üksikuid neist saab veebis huvi korral ise vabatahtlikult kasutada, nende kohta öeldakse riskikalkulaatorid (tabel 2).

Tabel 2. Eesnäärmevähi riskikalkulaatorid.

Viide	Nimi	Kirjeldus
THOMPSON jt, 2006	Prostate Cancer Prevention Trial Risk Calculator, PCPTRC, kaasaegselt tuntud kui ka PCPR	PSA, DRE, vanus, Aafrika-Ameerika rahvus, pere haiguslugu, varasema biopsia tulemus.
NAM jt, 2007	Sunnybrook prostate cancer risk calculator	Vanus, eesnäärme suurenemuse riskiskoor (IPSS), PSA tase, vaba/totaalne PSA, etniline päritolu, pere haiguslugu, DRE tulemus.
STEYERBERG jt, 2007 KRANSE jt, 2008 ROOBOL jt, 2012	Rotterdam Prostate Cancer Risk Calculator, RPCRC, kaasaegselt tuntud kui ka ERSPC või ERSPC-RC3/4,	Ilmselt kõige rohkem kirjeldatud ja täiendatud riskikalkulaator. Koosneb kuuest alamkalkulaatorist, sõltuvalt terviseandmete olemasolust. Suudab eristada kliiniliselt olulist ENV-d kõigist ENV juhtudest. Kasutab sisendina PSA, DRE ja TRUS tulemusi. Kasutab hindamisel Gleasoni skoori. RPCRC välised valideerimised on andud vastandlikke tulemusi kliinilise kasu kohta (PALSDOTTIR jt, 2023).
ANKERST jt, 2014	PCPTRC 2.0	Kõige esimese riskikalkulaatori edasiarendus. Eristatakse biopsia kolme võimalikku tulemust: negatiivne, lokaalne ENV, kaugelearenev ENV. Mudelisse lisati vaba/totaalse PSA suhe.
PARK jt, 2017	Korean Prostate Cancer Risk Calculator for High-Grade Prostate Cancer, KPCRC-HG*	Korea-spetsiifiline ENV riskikalkulaator. Sarnane Rotterdami kalkulaatorile.
ANKERST jt, 2018	PCBB	PCPTRC 2.0 edasiarendus.

* Ei ole veebis kättesaadav.

Ülaltoodud mudelite diskrimineerimisvõime näitaja (AUROC või C-indeks) jääb enamasti vahemikku 70...80%. Paraku need mudelid ei ole oma ebaselge kasu tõttu veel tavapärases praktikas kasutusel. Lisaks sellele nõuavad mitmed loodud riskimudelid uroloogiliste analüüside tulemusi (DRE, TRUS ja mõnes mudelis isegi varasem biopsia) ning üritavad seeläbi ennetada ebavajalikku biopsiat.

1.3.1. Geneetiline riskimudel

Lisaks eelpool (peatükis 1.2.2.2.) kirjeldatud üksikutele riskilookustele võib eesnäärmevähi kujunemisel osaleda kuni 5500 erinevat ühenukleotiidset polümorfismi ehk SNP-d üle kogu genoomi (ZHANG jt, 2020). 23. mai 2024 seisuga on NHGRI-EBI koostatud GWAS kataloogis dokumenteeritud pea 2800 SNP–ENV seost. Üks viis, kuidas sellist laia geneetilist riskiprofiili korraga arvesse võtta on arvutada igale mehele välja **polügeenne riskiskoor** ehk **PRS** (*polygenic risk score*, vahel ka **geneetiline riskiskoor** ehk **GRS**), mis ei vaata korraga ainult varasemalt väljatoodud üksikuid kandidaatgeene, vaid sadu erinevaid väikese mõjuga ühenukleotiidseid positsioone. Polügeenne riskiskoor on statistiline meetod, mis võimaldab

genotüpiseeritud ja imputeeritud markerite abil ennustada haigusriski. On kasutusel ka termin polügeenne skoor (*polygenic score*, PGS), mis on laiema tähendusega ja sobib ka mittepatogeensete fenotüüpide kirjeldamiseks (nt kehakaal, pikkus), kuid selles töös me hindame haigusriski ja seega eelistame terminit „polügeenne riskiskoor".

Erinevalt teistest riskikalkulaatoritest, kasutas S3M esmakordselt genotüübiandmeid (PRS-i ja ühte riskilookust eraldi), et hinnata geneetilist riski otse, mitte pere haigusloo kvalitatiivse küsimustiku abil kaude. Nüüdseks on tekkimas ka laiem arusaam, et polügeenseid riskiskoore võiks kasutada sõeluuringute suunamisel ja ennetusprogrammides, et eristada kliiniliselt olulisi vähijuhte, vähendades samal ajal suremust ja laiaulatusliku sõeluuringu (*blanket screening*) kulu ja kahjusid (HUNTLEY jt, 2023; JAMES jt, 2024).

PRS-ga haigusriski ennustamiseks kasutatakse genotüpiseeritud markereid. Markerid, mis tõenäoliselt osalevad haiguse kujunemisel, on suures osas välja selgitatud genoomiüleste seoseuuringute (*genome-wide association study*, GWAS) tulemusel. GWAS-st saadakse hinnang selle kohta, kui suur on iga vaadeldud alleeli suhteline mõju haiguse kujunemisele.

Formaalselt on PRS ühe indiviidi jaoks defineeritud kui riskialleelide mõjude kaalutud keskmine, kus riskialleelide mõjud on GWAS-st saadud regressioonikaalud (valem 1). (YANG jt, 2023)

$$\mathrm{PRS} = \sum_{i=1}^{n} \hat{\beta}_i \, \mathrm{SNP}_i = \hat{\beta}_1 \mathrm{SNP}_1 + \hat{\beta}_2 \mathrm{SNP}_2 + \ldots + \hat{\beta}_n \mathrm{SNP}_n, \qquad (1)$$

kus:

- SNP_i on i-nda lookuse riskialleelide arv ehk annus (*dosage*) indiviidi genoomis, teisisõnu kas indiviidi on lookuse suhtes homosügoot või heterosügoot (0...2);
- $\hat{\beta}_i$ on riskialleeli SNP_i suhteline mõju tunnuse avaldumisele (saadakse GWAS-st) (nn riskialleeli kaal, *weight*);
- i on konkreetse riskialleeli järjenumber;
- n on riskialleelide koguarv.

PRS on mudel, mis prognoosib riski ainult teadaolevate geneetiliste tegurite põhjal. Teadaolev tähendab, et PRS koostamisel kasutatavaid alleele on varasemalt kirjeldatud seoses ennustatava fenotüübiga, sealjuures on olemas varasematest statistilistest uuringutest saadud alleelide suhtelised mõjud riski suurenemisele.

Teaduskirjanduses on alates 2008. aastast kirjeldatud mitmeid ENV geneetilise riski mudeleid. Aja jooksul on riskimudelite keerukus suurenenud ja mudelid võtavad sisendina rohkem riskitegureid arvesse. Mõned varasemad geneetilise riski mudelid on tabelis 3.

Tabel 3. Valik konstrueeritud ENV geneetilise riski mudelitest. Lühendid: HPM – haruldane patogeenne mutatsioon, PH – pere haiguslugu (*family history*).

Viide	Meetod	Riskitegurid
ZHENG jt, 2008	Logistiline regressioon	Vanus, regioon, PH, 5 SNP.
XU jt, 2009	Logistiline regressioon	14 SNP-d, PH.
SUN jt, 2011	Korrutamine	SNP komplektid (5, 11 ja 28), PH.
MACINNIS jt, 2011	Kombinatsioon tõepärafunktsioonidest, Poissoni regressioon	26 levinud varianti, PH, retsessiivne HPM.
LINDSTRÖM jt, 2012	Logistiline regressioon	25 SNP-d, vanus, PH.
GRÖNBERG JT, 2015 STRÖM JT, 2018 PALSDOTTIR JT, 2023	Logistiline regressioon	Biomarkerid (PSA, hK2 jm), kliinilised tegurid, PRS_{254}, *HOXB13* riskivariant. Geneetilise riski lisamine riskiarvutusse parandas riskilahutust.
MARS jt, 2020	Coxi võrdelise riski mudel	PRS, vanus, PH, eesnäärmesuurenemuse ajalugu.
SHI jt, 2021	Coxi võrdelise riski mudel	18 HPM, PH, PRS.
COX ja COUPLAND, 2021	Cox võrdelise riski mudel	PSA tase, vanus, elatustase, etniline päritolu, suitsetamise staatus, rasked vaimsed haigused, diabeet, KMI, PH.
VARMA jt, 2023	Gradientvõimendus	Skriinimisandmed, PRS, kliinilised andmed.
NYBERG jt, 2023	Poissoni regressioon / Coxi võrdelise riski mudel	3 HPM, 1 oletuslik HPM, PRS_{268}, PH.

1.3.2. Olukord Eestis

Milline olukord on eesnäärmevähi personaalse riski hindamisega Eesti meestel? TASA jt (2020) viisid läbi hetkel ainsa teadaoleva Eesti-spetsiifilise eesnäärmevähi riskiuuringu, kus hinnati eestlaste geeniandmete põhjal viit erinevat polügeenset riskiskoori ja nende riskilahutusvõimet. Viie PRS-i seast leitud parim kalkulaator kasutas SCHUMACHER jt (2018) meta-analüüsi tulemil leitud 121 SNP-d ja nende kaalusid. Parima mudeli C-indeks oli 0,641 ja riskimäärade suhe (*hazard ratio*) oli 1,65. Ainuüksi PRS suutis eesnäärmeriski paremini eristada kui pere haigusloo teadmine, suutes tuvastada rohkem kui 3-kordset riskisuurenemist. Näiteks suurima geeniriskiga 41-aastasel mehel (0,95 riskikvantiil) on sama geenirisk, mis keskmisel 45-aastasel mehel, kuid 55-aastasena on suure geeniriskiga mehel sama geenirisk, mis keskmisel 68-

aastasel mehel. Teisisõnu tõestati, et eesnäärmevähi PRS on olulise kliinilise potentsiaaliga tööriist suure ENV-riskiga meeste eristamisel. Sellest ja EAU soovitustest lähtuvalt pakuti samas töös välja ka geeniriskist sõltuv PSA-skriinimise strateegia, mis kokkuvõttes soovitas suurema suhtelise geeniriskiga mehi varem ja tihedamini PSA suhtes mõõta.

Käesoleva töö autorile teadaolevalt on see ainuke Eestis kirjeldatud eesnäärmevähi riskipõhine valikumeetod Eesti meestel. Seni pole eestlaste jaoks veel loodud ühtegi kliinilist eesnäärmevähi riskimudelit, mis arvestab rohkem tegureid kui ainult PRS.

2. UURIMUS

2.1. Töö eesmärgid

Selle väitekirja eesmärgid on:

1) teada saada, kas kliiniline vaatlusteave suudab ennustada uusi eesnäärmevähijuhte Eesti meeste seas;
2) teada saada, kas eesnäärmevähi polügeenne riskiskoor koos kliinilise vaatlusteabega suudab prognoosida uusi eesnäärmevähijuhte paremini kui ainult kliiniline vaatlusteave;
3) hinnata, kui palju see potentsiaalne kliiniline riskimudel suudaks uusi juhte senisest praktikast paremini tuvastada.

2.2. Materjal ja metoodika

2.2.1. Andmed

2.2.1.1. RITA MAITT

RITA MAITT (masinõppe ja AI toega teenused) on populatsioonipõhine andmebaas, kus on 10% juhuvalitud Eesti isikukoodiga inimese terviseandmed, mis on esinesid aastatel 2012–2019 (OJA jt, 2023; EESTI TEADUSAGENTUUR, 2022). Kokku on andmebaasis 150 824 inimese terviseandmed. Andmed on teisendatud rahvusvahelisele OMOP CDM andmekujule (*Observational Medical Outcomes Partnership Common Data Model*) (OHDSI, 2021). OMOP CDM on meditsiiniliste vaatlusandmete organiseerimise standard. Sisuliselt see on relatsiooniline andmebaas, mis koosneb kümnetest tabelitest, mis organiseerivad teavet iga patsiendi haigusseisundite, ravimikasutuse, biomeetriliste mõõtmiste, tervishoiuvisiitide jm kohta. OMOP-i andmemudelis on kombineeritud kolm andmeallikat: digilugu (*electronic health records*), digiretseptid (*digital prescriptions*), Tervisekassa raviarved (*health insurance claims*).

2.2.1.2. Eesti geenivaramu

Tartu Ülikooli Eesti geenivaramu (edaspidi EGV) on populatsioonipõhine biopank, kus on ~210 000 inimese genotüübiandmed. Kõik EGV uuringualused on täitnud nõusolekuvormi ja

loovutanud vereproovi, millest eraldatud DNA põhjal on inimesi genotüpiseeritud. Inimeste terviseandmeid päritakse regulaarselt sellistest allikatest nagu Tervisekassa ja surmaregister (HAAN jt, 2024).

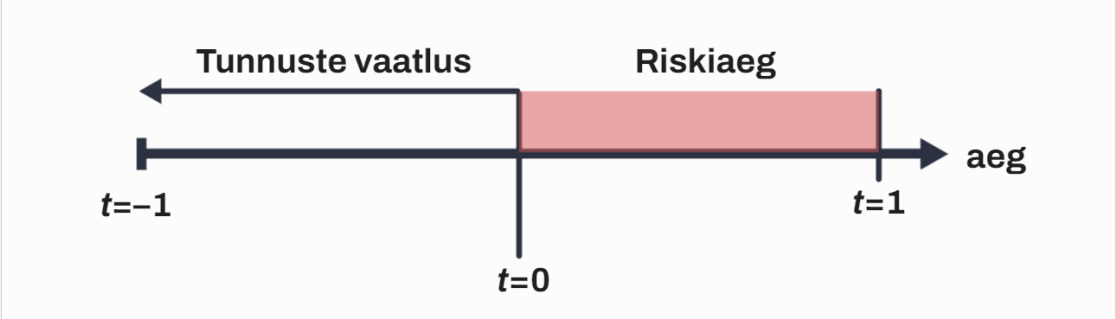
Selleks uurimistööks on antud Eesti bioeetika ja inimuuringute nõukogu luba nr 1.1-12/624.

2.2.2. Kliinilise riskimudeli arendamine

Kliinilise riskimudeli loomiseks kasutati OMOP andmekujule teisendatud kliinilisi vaatlusandmeid RITA MAITT andmebaasis. Mudeli treeningandmestiku koostamiseks kasutati OMOP andmestandardiga ühilduvat veebitarkvara ATLAS (OHDSI, 2024a). Mudeli treenimiseks ja valideerimiseks kasutati programmeerimiskeelt R ning OMOP andmetega ühilduvat R paketti PatientLevelPrediction v6.3.6 (REPS jt, 2018).

2.2.2.1. Uurimisprobleem

Riskimudel on statistiline mudel, millega ennustada, kas inimesel tekib kindla aja jooksul mingi uuritav tulem või mitte. Selline mudel võtab sisendiks mingid otseselt või kaudselt hinnatud **tunnused** (*features*, tähis *X*) ning hindab nende põhjal uuritavale inimesele tulemi (*outcome*) tekkimise tõenäosuse. Tunnused võivad olla näiteks diagnoosid, mis eelnevad ennustushetkele. Tunnuste asemel kasutatakse ka terminit mudeli **kovariandid** (*covariates*). Selles töös kasutatakse neid termineid läbisegi. Uuritav tulem selles uurimuses on eesnäärmevähk (lühidalt **ENV**). Riski mudeldamise skeem on visualiseeritud joonisel 2.



Joonis 2. $t = 0$ on ajahetk, mil ennustame mingile kindlalt määratud inimrühmale ehk **sihtkohordile** (*target cohort*, lühidalt T) tulemi tekkimise tõenäosust ehk haigusriski. Seejärel vaatleme, kes neist saavad mingi kindla vaatlusaja ehk **riskiaja** (*time-at-risk*, lühidalt TAR) jooksul uuritava tulemi (ENV). Me vaatleme tulemi tekkimist ajahetkeni $t = 1$. Tulemi tekkeriski ennustamiseks kasutame teavet kindlas vaatlusaknas, mis eelneb ennustushetkele $t = 0$. See vaatlusaken algab hetkest $t = -1$. Joonis on kohati ebatäpne, sest riskiaeg ei pea alati algama vahetult peale riskiprognoosi. Kõik näidatud ajahetked ja -intervallid sõltuvad uurimisprobleemist ja uurijast.

Uurimisprobleem: kes sihtkohordi liikmetest saavad riskiaja jooksul eesnäärmevähi? Selle probleemi lahendamiseks koostame ehk treenime mõned masinõppemudelid (*machine learning models*). Iga mudeli treenimiseks on vaja defineerida:

1) inimesed, kellele haigusriski tahame ennustada – sihtkohort;
2) inimesed, kes saavad vähi – tulemkohort (*outcome cohort*);
3) riskiaeg, mille jooksul vaatleme, kas inimesel tekib vähk või mitte ning mille lõpus saame hinnata, kas prognoos oli õige või vale;
4) tunnused, mille abil sihtkohordile vähki ennustame.

2.2.2.2. Uuringurahvastiku defineerimine

Mudeli treenimiseks on vaja treeningandmestikku. Tavaliselt on masinõppes sarnaste uurimisküsimuste puhul treeningandmestik sellisel kujul andmetabel, kus igas reas on üks inimene ja iga inimese jaoks on mitu tunnust (joonis 3).

Sugu	Vanus	X_1	X_2	...	X_n	Tulem
M	35	1	0		1	1
M	50	0	0		0	1
N	27	0	0		0	0
M	64	1	1		0	0
N	24	1	1		1	0
N	41	1	0		1	0
M	38	0	0		0	1

Joonis 3. Klassikalise väljanägemisega treeningandmestik, kus üks rida vastab ühele inimesele ja iga sinine tulp vastab ühele tunnusele, mida kasutatakse tulemi ennustamisel. Tunnused ja tulemid on esitatud indikaatortunnuste (*indicator variable*) kujul, kus „1“ tähendab tunnuse või tulemi olemasolu ja „0“ tähendab tunnuse või tulemi puudumist. Erandiks on mõned mõõte- või arvutustulemused, näiteks kehakaal või kehamassiindeks, kus indikaatortunnuse asemel on konkreetne väärtus.

Valmiskujul sellist andmetabelit veel pole. OMOP CDM-kujul andmebaasis on iga inimese terviseandmed jaotatud üle mitmete erinevate tabelite. See tähendab, et kõik tunnused ja tulemid on eraldi tabelites. Mudeli treenimiseks on tehnilistel põhjustel vaja siiski ideaalis ühte kompaktset andmetabelit nagu näidatud. Selleks on vaja defineerida sihtkohort ehk uuritavad inimesed, seejärel tulemkohort (inimesed kogu tervest andmestikust – ka väljapool sihtkohorti – kes said vähi), ja viimaks peab valima tunnused, mille abil sihtkohordile tulemit ennustada. Teades, kes on tulemkohordis, saame ühtlasi teada, kes sihtkohordist saavad suurima

tõenäosusega vähi. Tekkiv treeningandmestik on tuntud ka kui uuringurahvastik (*study population*).

OMOP CDM-kujule teisendatud terviseandmed on jaotatud kuueks kategooriaks: seisundid (*conditions*), ravimid (*drugs*), protseduurid (*procedures*), mõõtmised (*measurements*), vaatlused (*observations*) ja külastused (*visits*).

2.2.2.2.1. EESNÄÄRMEVÄHI MÕISTEHULK

Ühe haiguse diagnoosimiseks on üle maailma sageli võimalik valida mitme diagnoosisüsteemi ehk meditsiinisõnastiku vahel. Et leida andmebaasist üles kõikvõimalikud vaatlusandmed eesnäärmevähi kohta, on vaja koguda kokku kõik asjakohased diagnoosid. Üks tööriist selle jaoks OMOP CDM-s on mõistehulk (*concept set*), mis kogub kokku samatähenduslikud kliinilised faktid ehk mõisted (*concepts*), et standardiseerida meditsiiniliste sündmuste tähendust üle mitme erineva meditsiinisõnastiku. Enne, kui me saame andmebaasist välja võtta inimesed, kellel on olnud eesnäärmevähk (ja seda teadmist uurimuses edaspidi kasutada), on meil vaja defineerida, millised kliinilised vaatlusandmed vastavad eesnäärmevähile.

Üks rahvusvaheliselt levinud diagnoosisüsteem on RHK-10 (inglise keeles ICD-10) (WHO, 2004). RITA-MAITT andmebaasis RHK-10 koodid pole standardsed, mis tähendab, et nad ei seo endaga teiste sõnastike koode. OMOP-s on RHK-10 koodid enamasti teisendatud SNOMED koodideks. Ühe SNOMED koodi alla võib koonduda erinevate sõnastike kümneid samatähenduslikke koode. Selleks koostame mõistehulga, mille põhjal tuvastame andmebaasis eesnäärmevähile vastavad kliinilised sündmused (tabel 4). Teisisõnu: selle uurimuse raames on mehel eesnäärmevähk, kui tal leidub vähemalt üks alltoodud diagnoosidest. OMOP mõistekoodide otsimiseks kasutati veebiliidest ATHENA (OHDSI, 2024b).

Tabel 4. Eesnäärmevähi mõistehulk ehk mis vastab meie andmestikus eesnäärmevähile.

RHK-10 nimi	RHK-10	OMOP mõiste ID*
Eesnäärme pahaloomuline kasvaja	C61	200962
Eesnäärme kartsinoom *in situ*	D07.5	200970
Eesnäärme intraepiteliaalne neoplaasia	N42.31**	192681

* SNOMED kood; **RHK-10 kliiniline süsteem (ICD-10-CM)

2.2.2.2.2. SIHTKOHORT

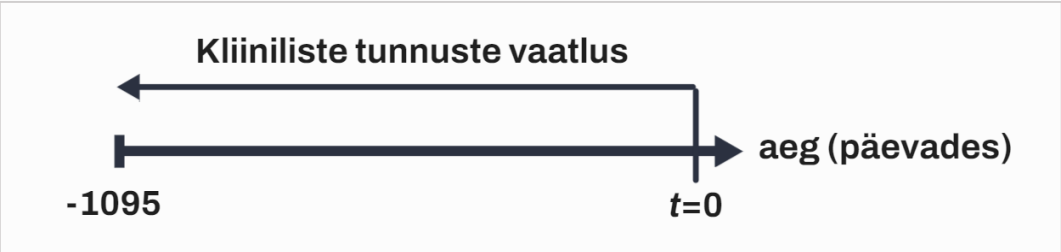
Sihtkohort on inimrühm ajahetkel $t = 0$, kellele prognoositakse tulemi tekkimise tõenäosust. Alternatiivselt saab sihtkohordi asemel öelda ka riskikohort. Kellele vähiriski ennustada? Neile,

kel pole seda varem olnud. Sihtkohorti ei kaasata neid, kellel juba on eesnäärmevähk (olnud). Samas, sihtkohort peaks olema maksimaalselt lai, et leida vähioht ka neil, kellel näiliselt ohtu pole. Mudelit pole mõtet treenida nendes gruppides , kelle puhul on teada, et nad kindlasti eesnäärmevähki ei saa (naised) või nendel, kellele on juba MRT või biopsia tehtud. Mudeli eesmärk on ikkagi võimalikult väheste vahenditega saada võimalikult täpne riskihinnang neile, kelle tervisekulgu esmapilgul ei osata kvalitatiivselt hinnata. Sihtkohordi definitsioon on toodud tabelis 5.

Tabel 5. Sihtkohordi definitsioon.

Parameeter	**Väärtus ja seletus**
Kohorti sisenemise sündmus	Inimene siseneb kohorti kohe, kui teda on andmestikus jälgitud vähemalt 1095 järjestikkust päeva (3 aastat). See on maksimaalne vaatlusperiood RITA MAITT andmestikus, peale mida on veel piisaval määral ka riskiaega (5 aastat). Kolmeaastase vaatluse põhjal koostatakse inimese riskitunnuste profiil ehk mudeli kovariandid ning tehakse kindlaks, et ta pole varasem vähihaige. Maksimaalne vaatlusperiood aitab välistada varasemad juhud, kes näiteks ühe aasta jooksul arstil ei käi ega kordusdiagnoosi ei saa.
Kaasamis-kriteeriumid	Kohorti jäävad ainult mehed.
	Vanus kohorti sisenemise hetkel: 40 a kuni 75 a. PUNAB (2023b) soovitas eesnäärmevähi kontrolliga alustada hiljemalt 40. eluaastast.
	Mehel pole mitte ühtegi varasemat eesnäärmevähi diagnoosi. Kui on, jäetakse see mees sihtkohordist välja. Siin kasutatakse eelnevalt defineeritud eesnäärmevähi mõistehulka, et tabada kõik juhud.
	Kolmeaastase vaatlusperioodi jooksul ei ole mitte ühtegi PSA-mõõtmist.
Kohordist väljumise sündmus	Puudub. Mees peab kohordis olema maksimaalselt kaua, et saaks tema vähiteket jälgida.

Sihtkohordi ajaloogika on kujutatud joonisel 4.



Joonis 4. Ajahetk $t = 0$ ehk indekskuupäev (*index date*) on kuupäev, mil inimene siseneb sihtkohorti ja millal mudel sellele inimesele riski prognoosib. Riski prognoosimiseks saab kasutada ainult neid andmeid, mis eelnevad indekskuupäevale. Tunnuste vaatlusperioodi põhjal sõelutakse ka välja need, kellel juba on eesnäärmevähk.

2.2.2.2.3. TULEMKOHORT

Tulemkohort on inimrühm tervest andmepopulatsioonist, kes ühel või teisel hetkel said uuritava tulemi (ENV). Tulemkohordis on kõik eesnäärmevähi sündmused populatsioonis, seega ka need inimesed, kes on väljaspool sihtkohorti (tabel 6).

Tabel 6. Tulemkohordi definitsioon.

Parameeter	**Väärtus ja seletus**
Kohorti sisenemise sündmus	Inimene siseneb kohorti kohe, kui tal tuvastatakse andmebaasis eesnäärmevähi diagnoos.
Kaasamiskriteeriumid	Kohorti jäävad ainult mehed.
Kohordist väljumise sündmus	Puudub.

2.2.2.2.4. TUNNUSTE VALIK

Et saada treeningandmed sellisele kujule, nagu on näidatud joonisel 3 (lk 20), kasutati R paketti FeatureExtraction (SCHUEMIE jt, 2024), mis konstrueerib igale sihtkohordi indiviidile kovariandid ja tulemid (*outcome labels*).

Kui anda masinõppemudelile ette hulk tunnuseid, siis mitmed masinõppemudelid suudavad valida nendest välja sellised tunnused, millel on ennustusjõud. Tasub teadvustada, et selles uurimuses loodava mudeli jaoks valitakse ennustavad tunnused välja suuresti statistilise assotsiatsiooni põhjal, mitte niivõrd põhjuslike seoste põhjal. Kuigi statistiline assotsiatsioon võib aidata mingite tunnuste põhjal kindlat tulemit ennustada, siis see ei pea olema tingimata põhjusliku seose pärast (RIJNBEEK, 2019). Sellegipoolest võivad ennustusjõudu demonstreerivad tunnused viidata sisulisele, bioloogilisele seosele. Mudeli treenimiseks antakse ette eesnäärmevähi-alasele kirjandusele tuginedes hulk tunnuseid, mille seast statistilised algoritmid valivad olulisemad.

Kliinilised vaatlused, mida kasutatakse mudeli treenimisel (tabel 7, järgmine lk). Tunnuste valik tehti kirjanduse ja kliinilis-empiirilise ekspertkogemuse põhjal (ptk 1.2.1, lk 9). Kuigi etniline päritolu ehk rass on oluline tegur eesnäärmevähi riski ennustamisel, siis kasutatavates andmetes ei ole rass kui tunnus eristatud, seega see jäetakse tunnuste valikust välja.

Tabel 7. Treeningandmestikku kaasatud tunnused. *n* on uuringus hõlmatud indiviidide arv aastatel 2012–2019.

RHK-10 nimi	RHK-10	OMOP mõiste ID*	*n*
Erektsioonihäire	F52.2	3655355	644
Alaseljavalu	M54	194133	26 660
Reievalu	M79.65	4302739	5513
Kõhu- ja vaagnapiirkonna valu	R10	200219	27 736
Kusemishäire	R30.0	197684	1375
Kusemispakitsus	R30.1	37205044	385
Veri uriinis	R31	437038	1725
Kusepidamatus	R32	197672	5024
Kusepeetus	R33	192450	1255
Anuuria ja oligouuria	R34	4101362	177
Polüuuria, sh öine kuselkäimine	R35	79936	145
Kusitieritis	R36	196821	11
Urineerimisega seotud muud raskused, sh vajadus kusemise alguses pingutada	R39.1	4010658	313
Eesnäärmesuurenemus	N40	197032	8701
Äge prostatiit	N41.0	193522	510
Krooniline prostatiit	N41.1	200445	3905
Eesnäärme-põiepõletik	N41.3	198807	52
Kehamassiindeks (kg/m^2)	–	4245997	72 946
Testosteroon	–	4020107	6201
Tubaka tarvitamine	–	4005823	571
Nikotiinisõltuvus	F17	4209423	430

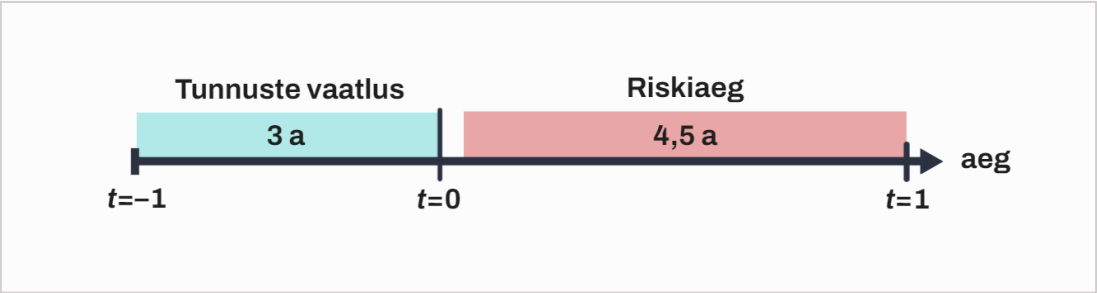
* SNOMED kood

Mudelisse valitakse selle treenimise käigus automaatselt ainult need tunnused, mis esinevad piisavalt paljude meeste vaatlusaknas. Ennustamisel ei saa tunnust kasutada, kui see: a) üldse ei esinenud enamiku indiviidide vaatlusaknas enne esmast ennustushetke või; b) seda esines liiga vähestel (<0,1%).

2.2.2.3. Populatsiooniseadistus

Mees on arvatud sihtkohorti ja talle ennustatakse nüüd vähiriski. Et prognoosida, kellel meestest vähk tekib ja et teada, kas mudel ennustas seda õigesti, on vaja määratleda, mis ajani vähi tekkimist vaadeldakse. Nüüd, kui kohordid on defineeritud, on vaja määrata ka riskiaeg

(*time-at-risk*), mil jälgitakse vähijuhtude teket ja mille lõpus saab hinnata mudeli ennustustäpsust (joonis 5).



Joonis 5. Riskimudeli ajaline loogika. Mudeli sisendiks on tunnuste vaatlusaknas kogutud teave erisuguste kliiniliste vaatluste kohta. Seda teavet kogutakse ajaintervallis $t = -1$ kuni $t = 0$, kus mudel prognoosib inimesele haigusriski. Selle töö raames on ajaintervalli pikkus 1095 päeva ehk 3 aasta. Riskiaeg algab 180 päeva pärast ennustushetke $t = 0$. Tulemil tekkimist vaadeldakse 4,5 aastat. Ajahetkeks $t = 1$ ehk riskiaja lõpuks on võimalik otsustada, kui täpselt mudel vähi teket prognoosis.

Kuna eesnäärmevähk on üpris aeglase kuluga haigus (mistõttupärast jäetakse mehed sageli ilma ravita jälgimisele), siis on seda parem, mida pikem on riskiaeg. Samas on vaja ka piisavalt pikka tunnuste vaatlusperioodi, et koguda võimalikult palju teavet mehe terviseseisundi kohta ja et välistada aktiivseid vähijuhte.

Inimesed, kes mingil põhjusel treeningandmestikust kaovad enne riskiaja lõppu (nt surevad muudel põhjustel), jäetakse treeningandmestikust välja.

2.2.2.4. Algoritmi valik

Kliinilise mudeli konstrueerimiseks on valida mitme masinõppealgoritmi vahel. Kasutati kolme konkureerivat algoritmi: logistiline regressioon koos lassoregulariseerimisega, gradientvõimendus (*gradient boosting*) ja otsustusmets (*random forest*).

2.2.2.5. Mudeli hüperparameetrid

Mudeli parimate hüperparameetrite jaoks kasutati võreotsingut (*grid search*), mis omakorda kasutas viiekordset ristvalideerimist (*5-fold cross-validation*).

2.2.2.6. Mudeli sobitamine

Mudeli sobitamine (*fitting*) on protsess, kus arvutustarkvara leiab tunnustele ehk mudeli kovariantidele regressioonikaalude hinnangud $\hat{\beta}$ (mitte segi ajada PRS arvutuse kaaludega),

kasutades selleks eelnevalt koostatud treeningandmestikku. Et aga mudeli ennustusvõimet koheselt hinnata, tuleb tekitatud treeningandmestik poolitada kaheks, kasutades 80/20 suhet:

- 80% andmetest kasutatakse mudeli parameetrite sobitamiseks ja optimeerimiseks (võreotsing leiab parimad hüperparameetrid, kasutades viiekordset ristvalideerimist);
- 20% andmetest kasutatakse mudeli ennustusvõime esmaseks hinnanguks. Need andmed pannakse mudeli sobitamise ajaks kõrvale (n-ö *hold-out* andmestik).

2.2.3. Polügeense riskiskoori arvutamine

Polügeense riskiskoori arvutamiseks EGV uuritavatele kasutati meetodit PRS-CS-auto (GE jt, 2019). Variantide mõjuhinnanguks $\hat{\beta}$ (võrrand 1, lk 15) kasutati SCHUMACHER jt (2018) meta-analüüsis tuletatud alleelide mõjusuurusi. Et kohendada neid mõjuhinnanguid populatsiooni geneetilise struktuuri suhtes, kasutati 1000 Genomes Project *phase 3* Euroopa LD referentspaneeli. PRS-CS hindas igale variandile uue, kohandatud mõju (*posterior effect size*) ning arvutas seejärel igale uuringaluseleuuritavale polügeense riskiskoori. Polügeensed riskiskoorid standardiseeriti, st PRS-de jaotus teisendati standardse normaaljaotuse kujule, $PRS \sim N(0,1)$.

2.2.4. Mudelite ennustusvõime hindamine

Statistilise mudeli arendamise käigus treenitakse korraga mitmeid mudeleid. Kõige lihtsam viis on kasutada eksimismaatriksit (*confusion matrix*) (tabel 8)

Tabel 8. Eksimismaatriks (LEVER jt, 2016).

		Mis on tegelik väärtus?	
		P	N
Mida ennustas mudel?	P	Õigepositiivne (P^+)	Valepositiivne (P^-)
	N	Valenegatiivne (N^-)	Õigenegatiivne (N^+)

Siinse uurimuse kontekstis on valenegatiivne (N^-) kõige halvem tulemus – mees, kellele mudel ei ennustanud vähki, siiski sai vähi. Valepositiivne (P^-) ei ole sama kahjulik, sest mudeli

kliinilisel rakendamisel tähendaks see rohkemate meeste suunamist arstlikku jälgimisse. See võib nõuda vajalikust rohkem ressurssi, kuid ei ole sama kaaluga eksimus, kui valenegatiivne. Mudeli ennustusvõime hindamiseks (*model evaluation*) ja parima mudeli valimiseks on kasutada mitmeid statistikuid. Mõned neist on toodud tabelis 9.

Tabel 9. Mudeli ennustusvõime näitajad (STEYERBERG, 2019; ÇORBACIOĞLU ja AKSEL, 2023). Ennustusvõime näitajad on paigutatud selle uurimistöö raamistikku, et toetada selle töö tulemuste tõlgendamist. Kõik näidatud statistikud tuginevad eksimismaatriksile.

Statistik	**Seletus**
Tundlikkus (*sensitivity*)	Tõenäosus, et mudel annab positiivse hinnangu, kui mees on tegelik vähijuht (ENV^+).
Spetsiifilisus (*specificity*)	Tõenäosus, et mudel annab negatiivse hinnangu, kui mees on tegelikult vähivaba (ENV^-).
AUROC (*area under ROC curve*), samuti tuntud kui AUC	Tõenäosus, et mudel hindab vähihaigele mehele (ENV^+) suuremat vähitõenäosust kui juhuslikule vähivabale (ENV^-) mehele. $AUROC = 0$ korral on kõik mudeli ennustused valed. $AUROC = 0{,}5$ korral hindab mudel vähiriski sama hästi kui mündivise. $AUROC = 1$ korral on mudeli ennustused alati õiged. „Mõõdukalt hea“ mudeli AUROC on sageli umbes 0,75 ja „hea“ mudeli AUROC on vähemalt 0,8 (HOND jt, 2022).
AUPRC (*area under precision-recall curve*)	Võib aidata ennustusvõimet hinnata olukorras, kus vähijuhte on kontrollidega võrreldes eriti vähe.

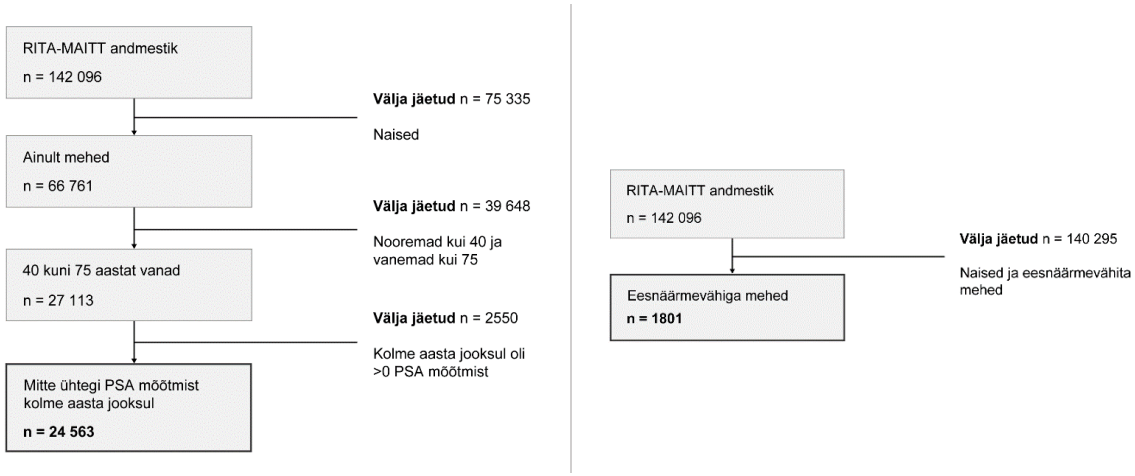
Olukorras, kus kahe mudeli AUROC ja AUPRC on sarnased ning valenegatiivsete hind (tuvastamata jäänud vähijuhud) on kõrgem kui valepositiivsete hind (vähivabad mehed, kellel tuvastati vähk), soovitab MCDERMOTT jt (2024) kasutada optimaalse mudeli valikul AUROC-i, sest see tagab konkreetses olukorras parema valenegatiivsete vähendamise.

2.3. Tulemused

2.3.1. Kliiniline riskimudel

2.3.1.1. Andmestik

Uuringukohordi valiku loogiline diagramm on näidatud joonisel 6.



Joonis 6. Sihtkohort (vasakul) ja tulemkohort (paremal). RITA-MAITT andmestik hõlmab endas terviseandmeid ajavahemikus 2012 kuni 2019.

Treeningandmestik moodustatakse nende kahe kohordi põhjal, võttes arvesse riskimudeli ajalist loogikat: 3 aastat tunnuste vaatlust ja 4,5 aastat riskiaega. Tekkivas 40–75-aastaste treeningandmestikus on 24 563 meest, kellest 772 saab nelja-aastase riskiaja jooksul eesnäärmevähi. Lisaks sellele tehti treeningandmestikust kaks alamandmestikku: 40–59-aastased (16 506 meest, kellest 153 on vähijuhud) ning 60–75-aastased (8057 meest, kellest 619 on vähijuhud).

2.3.1.2. Tunnuste automaatne valik

Mudeleid koostati kõigi kolme vanusegrupi jaoks. Tunnused valiti ja kombineeriti tabel 7 nimekirja põhjal (lk 24) toodud kovariandid. Kõikides valideerimistsüklites osutus, et lasso logistiline regressioon oli suurima ennustustäpsusega algoritm. Lassoregulariseerimise tulemil jäeti välja järgmised kovariandid, sest need omandasid nulliga võrduvad regressioonikordajad:

- alumiste kuseteede sümptomid (12 diagnoosi);
- erektsioonihäire;
- luu-, reie- ja seljavalud (9 diagnoosi);

- eesnäärme-põiepõletik;
- alkoholi kuritarvitamine;
- nikotiinisõltuvus (2 diagnoosi).

See tähendab, et ülaltoodud tunnused ei omanud siinses uuringus mitte mingit ennustusjõudu. Tunnused, mis jäid automaatse valiku järel alles:

- eesnäärmesuurenemus;
- krooniline prostatiit;
- kehamassiindeks.

2.3.1.3. Ennustusvõime RITA MAITT andmestikus

Mudeli treeniti järgmistele vanusegruppidele: 40–75, 40–59 ja 60–75. Kliinilise riskimudeli ennustusvõime hindamiseks kasutati treeningandmestikust 20% väljajäetud andmeid. Parima ennustusvõimega mudelid on toodud tabelis 10. Teistes vanusegruppides treenitud mudelid olid nõrgema ennustusvõimega.

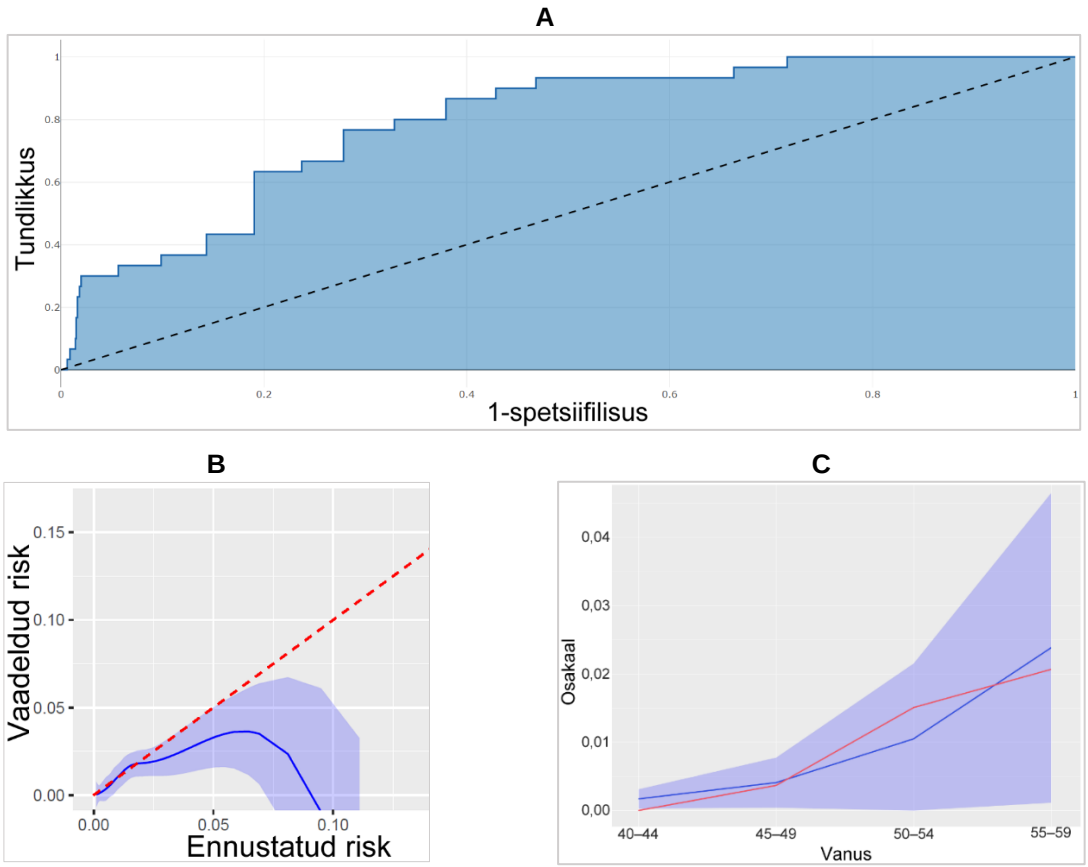
Tabel 10. Kliiniliste mudelite ennustusvõime vanusegrupis 40–59 (16 506 meest, kellest 153 olid juhud). Mudel ennustab 5 aasta riski.

Kovariandid	AUROC (95% CI)	AUPRC
ES	0,6622 (57,45–75,00)	3,72
KP	0,5502 (48,82–61,21)	1,47
Vanus	0,7551 (69,51–81,52)	1,85
Vanus, ES	0,8067 (74,17–87,16)	4,19
Vanus, KP	0,7751 (71,44–83,59)	2,20
Vanus, ES, KP	**0,8125 (74,62–87,89)**	**4,41**
Vanus, ES, KP, KMI	0,8080 (73,80–87,80)	4,26

ES – eesnäärmesuurenemus; KP – krooniline prostatiit; KMI – kehamassiindeks

Parim mudel meie uuringus on seega eelviimane mudel, millel on suurim AUROC ja AUPRC. Mudel kasutab eesnäärmevähi riski ennustamiseks kolme tunnust: vanus, eesnäärmesuurenemus ja krooniline prostatiit. Edaspidi nimetatakse seda kui VESP-mudel (**v**anus, **e**esnäärme**s**uurenemus, **p**rostatiit).

VESP-mudeli ennustusvõime graafikud on toodud joonisel 7.



Joonis 7. Kliinilise mudeli VESP ennustusvõime. (**A**) AUROC. (**B**) Sile kalibreerimiskõver. Sinine joon tähistab Loessi algoritmiga silutud kalibreerimiskõverat, punane tähistab ideaalset kalibratsiooni. VESP-mudel mõõdukalt ülehindab riski, sest VESP-i kalibreerimiskõver jääb allapoole diagonaali. (**C**) Eesnäärmevähi diagnoosi esinemuse ülevaade vanusegruppides. Sinine joon on mudeli ennustatud risk, helesinine ala on ennustatud riski 95% usaldusvahemik, punane joon on vaadeldud risk. Graafikult on näha, et loodud mudel on vanuste lõikes üpris hästi kalibreeritud.

2.3.1.4. Parima mudeli kirjeldus

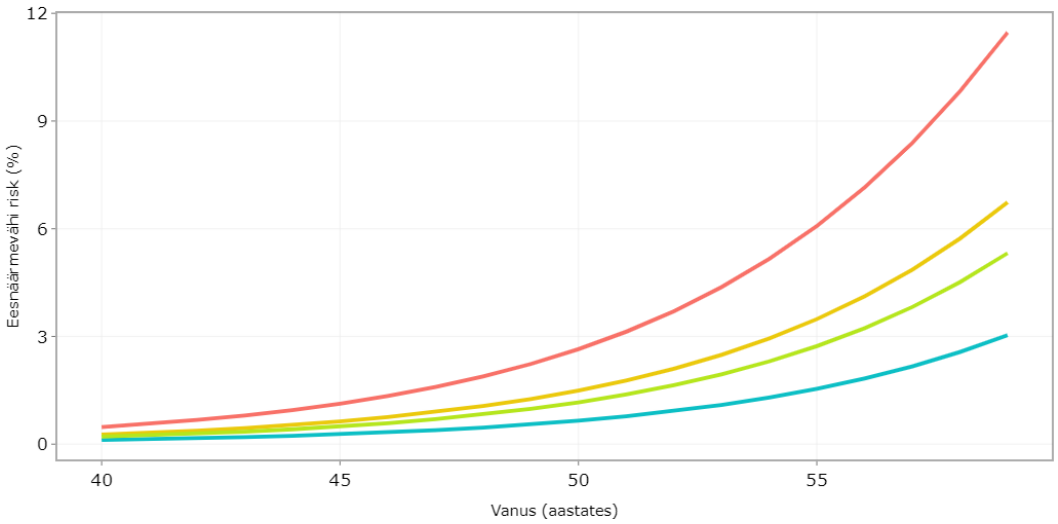
Parima ennustusvõimega mudeli 40–59 aastaste meeste jaoks ehk VESP-mudeli ($AUROC = 0{,}813$, $AUPRC = 4{,}41$) regressioonikaalud on esitatud tabelis 11.

Tabel 11. Eesnäärmevähi kliinilise riskimudeli (VESP) parameetrid. Mudeli vabaliige $\widehat{\beta_0} = -13{,}7$.

Argumenttunnus	$\widehat{\boldsymbol{\beta}}$	**OR****
Eesnäärmesuurenemuse esinemine* 3 vaatlusaasta jooksul	0,8353	2,31
Kroonilise prostatiidi esinemine* 3 vaatlusaasta jooksul	0,5842	1,79
Vanus aastates	0,1735	1,19

* Kodeeritud indikaatortunnusena: 1 – „esines“, 0 – „ei esinenud“; ** $OR = e^{\beta}$ näitab, kui mitu korda on haiguse olemasolul (või vanuse puhul iga täiendava eluaastaga) vähirisk suurem.

Eesnäärmevähi kumulatiivne risk on toodud joonisel 8.



Joonis 8. Eesnäärmevähi kumulatiivse riski graafik. Neli kõverat kirjeldavad nelja erinevat meest: **sinine** kirjeldab meest, kellel pole ei eesnäärmesuurenemust (ES^-) ega kroonilist prostatiiti (KP^-). **Roheline** kirjeldab meest, kellel on krooniline prostatiit (KP^+), kuid ei ole eesnäärmesuurenemust (ES^-). **Kollane** kõver on (ES^+, KP^-) mees. **Punane** kõver on (ES^+, KP^+) mees. Jooniselt on näha, et mida rohkem eesnäärmeprobleeme, seda suurem on ka eesnäärmevähi risk.

2.3.1.5. Ennustusvõime EGV andmestikus

TÜ arvutiteaduse instituut haldab ka OMOP kujule teisendatud Eesti geenivaramu uuritavate terviseandmeid (praegu arendusjärgus). Seda andmestikku ei olnud võimalik mudeli treenimiseks kasutada tehniliste piirangute tõttu, kuid seda on võimalik kasutada olemasolevate mudelite valideerimiseks (tabel 12).

Tabel 12. Kliiniliste mudelite ennustusvõime EGV OMOP andmetel, vanusegrupis 35–88 (30 045 meest, kellest 622 olid vähijuhud). Mudel ennustab 5-aastast riski.

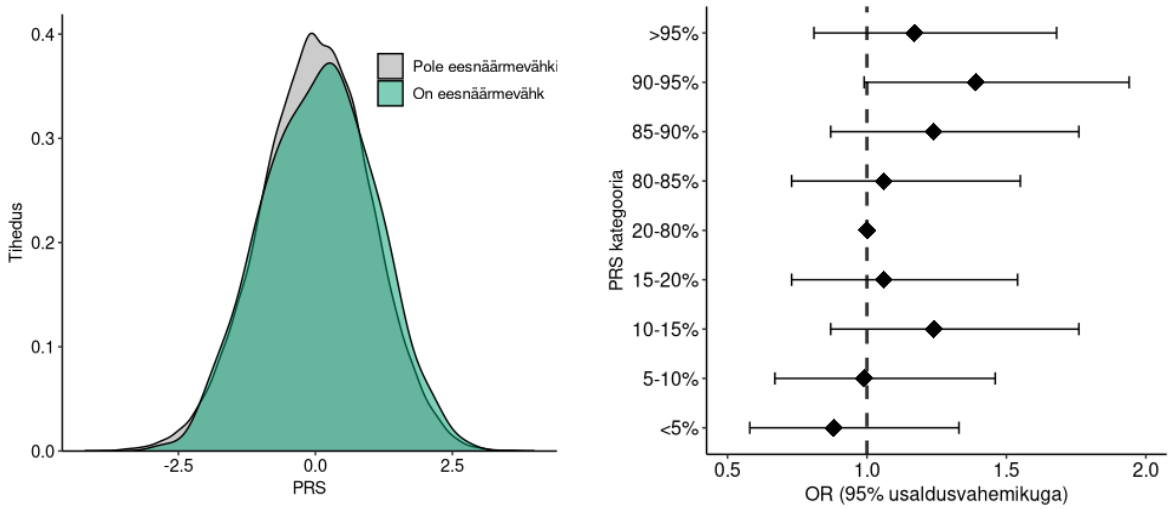
Kovariandid	AUROC (95% CI)	AUPRC
ES	0,6249 (0,6061–0,6437)	4,11
KP	0,5118 (0,5022–0,5214)	2,14
Vanus	0,8473 (0,8366–0,8579)	7,68
Vanus, ES	0,8514 (0,8408–0,8619)	7,80
Vanus, KP	0,8484 (0,8378–0,8590)	7,74
Vanus, ES, KP (VESP-mudel)	**0,8515 (0,8409–0,8621)**	**7,83**
Vanus, ES, KP, KMI	0,8515 (0,8410–0,8621)	7,83

ES – eesnäärmesuurenemus; KP – krooniline prostatiit; KMI – kehamassiindeks

Osutub, et laiemas vanuserühmas on mudeli ennustusvõime isegi parem.

2.3.2. PRS

Polügeensete riskiskooride jaotus on joonisel 9.



Joonis 9. (**A**) Polügeensete riskiskooride jaotused 35–88-aastaste eesnäärmevähiga meeste seas (roheline) ja eesnäärmevähita meeste seas (hall). (**B**) Polügeense riski kategooria ja vastav eesnäärmevähi risk 35–88-aastaste seas, mõõdetuna keskmise polügeense riski suhtes.

Ilma eesnäärmevähita meeste (n-ö kontrollgrupi) keskmine eesnäärmevähi-PRS oli 0,0014 ning eesnäärmevähiga meeste keskmine eesnäärmevähi-PRS oli 0,07. Kahe keskväärtuse erinevuse võrdlemiseks ($H_0: \mu_{ENV+} = \mu_{ENV-}$) teostati Welchi t-test, mille tulemil saadi p-väärtus = 0,1. See tähendab, et ENV$^+$ ja ENV$^-$ meeste PRS-de keskväärtused ei erine statistiliselt oluliselt (olulisuse nivoo $\alpha = 0{,}05$). Kuna PRS ei oma piisavat ennustusjõudu, siis selle lisamine mudelisse teeb mudelit põhjendamatult keerulisemaks.

2.4. Arutelu

Käesolevas töös arendati välja riskimudel eesnäärmevähi riski ennustamiseks, kasutades selleks kliinilisi vaatlusandmeid. Osutus, et mudeli statistiliselt kõige olulisemad argumenttunnused olid vanus, eesnäärme suurenemine ja krooniline prostatiit. Kuigi statistiline seos siin mudelis ei viita bioloogilisele põhjuslikkusele, saab seda teavet siiski riski prognoosimisel kasutada. Parima mudeli ehk VESP-mudeli ennustusvõimet peegeldav AUROC oli 0,852. VESP-mudeli kohaselt on 50-aastasel mehel, kellel on diagnoositud eesnäärmesuurenemus ja kroonililine prostatiit sama suur risk, kui mehel, kes on 59 aastat vana ja kellel ei ole kumbagi eesnäärme haigusseisundit (vastavalt ~2,7% ja 3%). Nii eesnäärmesuurenemusega kui ka kroonilise

prostatiidiga mehel on 59-aastaselt ligi 12% viie aasta haigestumise tõenäosus. Selles uurimuses PRS ei suutnud üheski vanusegrupis riski eristada. Kuna riskimudelid peaksid olema võimalikult lihtsad, et võimalikult väheste vahenditega riski ennustada, siis PRS-i lisamine riskiarvutusse ei õigusta selle ainelist kulu ja vähest potentsiaalset kasu.

Uurimusel olid mitmeid piiranguid. Järgmised uurimused võiksid neid puudujääke võimaluse korral arvestada.

1. On vaja suuremat treeningandmestikku. Kui osutub tehniliselt võimalikuks, siis tasub mudel konstrueerida EGV OMOP andmestikul ja valideerida RITA MAITT andmetel.
2. Mudelis ei eristatud agressiivset eesnäärmevähki mitteagressiivsest. Eesnäärmevähi ravis oleks hea teada, kas vähil on suur tõenäosus kujuneda metastaatiliseks.
3. Mudelisse võiks kaasata võimaluse korral rohkemaid riskitegureid, nagu etniline päritolu või rass, eesnäärme ruumala, eesnäärmesekreedi analüüsi tulemused. Samuti võivad rolli mängida välistegurid (näiteks kliima, kus mees elab: soe ja kuiv kliima võib olla eesnäärmele parem kui külm ja niiske) (ST-HILAIRE jt, 2010). Lisaks PSA mõõtmisele tasuks kasutada ka PSA-D näitu (*PSA density*). See on suhe PSA taseme ja eesnäärme ruumala vahel. Mida suurem PSA-D, seda tõenäolisem on, et vähi olemasolul on kliiniliselt oluline (YUSIM jt, 2020). Samuti võivad rolli mängida metaboliidid (ZADRA ja LODA, 2018). VICKERS jt (2023) tõid välja ka sotsiaalmajandusliku asümmeetria olulisuse eesnäärmevähi levimuses.
4. Eesnäärmevähi varajases ennetuses on lootustandev lahendus piltdiagnostika, eeskätt MRT. Riskipõhine lähenemine ja piltdiagnostika on uus värske uurimissuund.

Loodud mudel võib aidata kliinilises olukorras eristada indiviidide riski, kasutades ainult teavet vanuse, eesnäärmesuurenemuse ja kroonilise prostatiidi kohta. Olukorras, kus Eestis on tõenäoliselt hinnanguliselt 10 000 aktiivset eesnäärmevähijuhtu (TAI, 2023), võib loodud mudel aidata tuvastada hinnanguliselt kuni 1000 täiendavat juhtu, mis muidu jääksid vanusevahemikus 40 kuni 59 tuvastamata.

Kokkuvõte

Selles töös sooviti teada saada, kas kliiniline vaatlusteave suudab ennustada uusi eesnäärmevähijuhte Eesti meeste seas. Selleks kasutati RITA MAITT andmestikku, mis sisaldab endas terviseandmeid 10% Eesti isikukoodiga kodanike juhuvalimist.

Koostati mitmeid mudeleid, millest parimaks osutus VESP-mudel (**v**anus, **e**esnäärme**s**uurenemus, krooniline **p**rostatiit). Selle mudeli $AUROC$ viie aasta riski hindamisel oli 0,852. Järeldati, et kliinilised vaatlusandmed suudavad eesnäärmevähi riski tähendusväärselt eristada. VESP-mudeli kohaselt on 50-aastasel mehel, kellel on diagnoositud eesnäärmesuurenemus ja kroonililine prostatiit sama suur risk, kui mehel, kes on 59 aastat vana ja kellel ei ole kumbagi eesnäärme haigusseisundit (vastavalt ~2,7% ja 3%). Nii eesnäärmesuurenemusega kui ka kroonilise prostatiidiga mehel on 59-aastaselt ligi 12% tõenäosus haigestuda 5 aasta jooksul.

Et hinnata polügeensete mõjude ennustusvõimet, koostati Eesti geenivaramu andmestiku põhjal eesnäärmevähi polügeenne riskiskoor. PRS ei suutnud riski eristada. Selle riskilahutusvõime oli nullilähedane, mistõttu ei olnud õigustatav ka PRS-i lisamine mudelisse. VESP-mudelil on perspektiiv tuvastada Eestis kuni 1000 täiendavat uut eesnäärmevähi juhtu, mis muidu jääksid vanusevahemikus 40 kuni 59 tuvastamata.

Summary

Development of a prostate cancer risk model on Estonian health data

Telver Objärtel

Summary

Prostate cancer (PCa) is the leading cancer in men. It is the second-biggest cause of cancer mortality in men, right after lung cancer. It is notoriously difficult to detect, as it can remain asymptomatic for several years. There is a lack of nationally mandated PCa screening programmes. In the light of this shortcoming, we aimed to develop a PCa risk model that uses clinical observational data to predict incident PCa in men starting from the age of 40. We used population-based health data from RITA MAITT ($n = 150k$) to train and internally validate a PCa risk model. The model was validated on a different dataset, the Estonian Genome Centre (EGCUT) study ($n = 210k$). The best model has an $AUROC$ of 0,852, demonstrating the ability to differentiate risk.

We also calculated a polygenic risk score for the EGCUT male participants in the age group of 35 to 88 years. The risk discrimination of the PRS was near zero. Since a risk model should be as parsimonious as possible and since the PRS didn't predict the risk at all, adding PRS to the clinical risk calculation does not justify the cost and utility of it.

The developed model has the potential to identify up to additional 1000 men on a population scale.

Kasutatud kirjandus

ALADWANI, M., LOPHATANANON, A., OLLIER, W., MUIR, K. (2020). Prediction models for prostate cancer to be used in the primary care setting: a systematic review. *BMJ Open*, *10*(7). https://doi.org/10.1136/bmjopen-2019-034661

ANKERST, D. P., HOEFLER, J., BOCK, S., ..., THOMPSON, I. M. (2015). Prostate cancer prevention trial risk calculator 2.0 for the prediction of low- versus high-grade prostate cancer. *Urology*, *83*(6). https://doi.org/10.1016/j.urology.2014.02.035

ANKERST, D. P., STRAUBINGER, J., SELIG, K., ..., VICKERS, A. J. (2018). A Contemporary Prostate Biopsy Risk Calculator Based on Multiple Heterogeneous Cohorts. *Eur Urol.*, *74*(2). https://doi.org/10.1016/j.eururo.2018.05.003

AMERICAN CANCER SOCIETY. (2019). Cancer Facts & Figures for African Americans 2019-2021. *American Cancer Society*.

BMJ. (2024). TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. https://doi.org/10.1136/bmj.q902

CLARK, T. G., BRADBURN, M. J., LOVE, S. B., ALTMAN, D. G. (2003). Survival Analysis Part I: Basic concepts and first analyses. Br J Cancer, 89(2): 232– https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2394262/

COX, J., COUPLAND, C. (2021). Predicting the risk of prostate cancer in asymptomatic men: a cohort study to develop and validate a novel algorithm. *Br J Gen Pract.*, *71*(706).

ÇORBACIOĞLU, Ş. K., AKSEL, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*, *23*(4). https://doi.org/10.4103/tjem.tjem_182_23

EUROPEAN ASSOCIATION OF UROLOGY. (2024). EAU-EANM-ESTRO-ESUR-ISUP-SIOG Guidelines on Prostate Cancer.

GE, T., CHEN, C.-Y., NI, Y., FENG, Y.-C. A., SMOLLER, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, 10. https://doi.org/10.1038/s41467-019-09718-5

GRÖNBERG, H., ADOLFSSON, J., ALY, M., ..., EKLUND, M. (2015). Prostate cancer screening in men aged 50-69 years (STHLM3): a prospective population-based diagnostic study. *Lancet Oncol.*, *16*(16).

HAAN, E., KREBS, K., VÕSA, U., BRIKELL, I., LARSSON, H., LEHTO, K. (2024). Associations between attention-deficit hyperactivity disorder genetic liability and ICD-10 medical conditions in adults: utilizing electronic health records in a Phenome-Wide Association Study. *Psychological Medicine*. https://doi.org/10.1017/S0033291724000606

HJELMBORG, J. B., SCHEIKE, T., HOLST, K., ..., MUCCI, L. A. (2014). The Heritability of Prostate Cancer in the Nordic Twin Study of Cancer. *Cancer Epidemiol Biomarkers Prev*, *23*(11). https://doi.org/10.1158/1055-9965.EPI-13-0568

HOND, A. A. H., STEYERBERG, E. W., CALSTER, B. (2022). Interpreting area under the receiver operating characteristic curve. *The Lancet Digital Health*, *4*(12). https://doi.org/10.1016/S2589-7500(22)00188-1

HUNTLEY, C., TORR, B., SUD, A., ..., TURNBULL, C. (2023). Utility of polygenic risk scores in UK cancer screening: a modelling analysis. *The Lancet Oncology*, *24*(6). https://doi.org/10.1016/S1470-2045(23)00156-0

ILIC, D., DJULBEGOVIC, M., JUNG, J. H., HWANG, E. C., ZHOU, Q., CLEVES, A., AGORITSAS, T., DAHM, P. (2018). Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. *BMJ*, 362. https://doi.org/10.1136/bmj.k3519

JAMES, N. D., TANNOCK, I., N'DOW, J., FENG, F., GILLESSEN, S., ALI, S. A. (2024). The Lancet Commission on prostate cancer: planning for the surge in cases. *The Lancet*, *403*(10437). https://doi.org/10.1016/S0140-6736(24)00651-2

KRANSE, R., ROOBOL, M., SCHRÖDER, F. H. (2008). A graphical device to represent the outcomes of a logistic regression analysis. *Prostate*, *68*(15). https://doi.org/10.1002/pros.20840

LEVER, J., KRZYWINSKI, M., ALTMAN, N. (2016). Classification evaluation. *Nature Methods*, *13*. https://doi.org/10.1038/nmeth.3945

LINDSTRÖM, S., SCHUMACHER, F. R., COX, D., ..., KRAFT, P. (2012). Common genetic variants in prostate cancer risk prediction – Results from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3). *Cancer Epidemiol Biomarkers Prev*, *21*(3).

MACINNIS, R. J., ANTONIOU, A., EELES, R. A., ..., EASTON, D. F. (2011). A risk prediction algorithm based on family history and common genetic variants: application to prostate cancer with potential clinical impact. *Genet Epidemiol.*, 35(6).

MARS, N., KOSKELA, J. T., RIPATTI, P., ..., RIPATTI, S. (2020). Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nature Medicine*, *26*.

MARTIN, R. M., TURNER, E. L., YOUNG, G. J., ..., DONOVAN, J. L. (2024). Prostate-Specific Antigen Screening and 15-Year Prostate Cancer Mortality: A Secondary Analysis of the CAP Randomized Clinical Trial. *JAMA*, *331*(17), 1460–1470. https://doi.org/10.1001/jama.2024.4011

MCDERMOTT, M., HANSEN, L. H., ZHANG, H., ANGELOTTI, G., GALLIFANT, J. (2024). A Closer Look at AUROC and AUPRC under Class Imbalance. *arXiv*. https://doi.org/10.48550/arXiv.2401.06091

MCNALLY, C. J., RUDDOCK, M. W., MOORE, T., MCKENNA, D. J. (2020). Biomarkers That Differentiate Benign Prostatic Hyperplasia from Prostate Cancer: A Literature Review. *Cancer Manag Res*, *12*, 5225–5241. https://doi.org/10.2147/CMAR.S250829

MÜLLER, E., MEISSNER, V. H., KRON, M., SCHIELE, S., SCHULWITZ, H., GSCHWEND, J. E., HERKOMMER, K. (2022). Prostate-specific antigen levels depending on frequency of ejaculation and time since last ejaculation. *The Journal of Sexual Medicine*, *19*(11). https://doi.org/10.1016/j.jsxm.2022.08.173

NAM, R. K., TOI, A., KLOTZ, L. H., ..., KATTAN, M. W. (2007). Assessing Individual Risk for Prostate Cancer. *Journal of Clinical Oncology*, *25*(24). https://doi.org/10.1200/JCO.2007.10.6450

NYBERG, T., BROOK, M. N., FICORELLA, L., ..., ANTONIOU, A. C. (2023). CanRisk-Prostate: A Comprehensive, Externally Validated Risk Model for the Prediction of Future Prostate Cancer. *J. Clin Oncol*, *41*(5). https://doi.org/10.1200/JCO.22.01453

OJA, M., TAMM, S., MOOSES, K., ..., REISBERG, S. (2023). Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned. *JAMIA Open*, *6*(4). https://doi.org/10.1093/jamiaopen/ooad100

PALSDOTTIR, T., GRÖNBERG, H., HILMISSON, A., EKLUND, M., NORDSTRÖM, T., VIGNESWARAN, H. T. (2023). External Validation of the Rotterdam Prostate Cancer Risk Calculator and Comparison with Stockholm3 for Prostate Cancer Diagnosis in a Swedish Population-based Screening Cohort. *European Urology Focus*, *9*(3). https://doi.org/10.1016/j.euf.2022.11.021

PARK, J. Y., YOON, S., PARK, M. S., ..., BYUN, S.-S. (2017). Development and External Validation of the Korean Prostate Cancer Risk Calculator for High-Grade Prostate Cancer: Comparison with Two Western Risk Calculators in an Asian Cohort. *PLoS One*, *12*(1). https://doi.org/10.1371/journal.pone.0168917

RAI, S., MISHRA, P., GHOSHAL, U. D. (2021). Survival analysis: A primer for the clinician scientists. *Indian J Gastroenterol*, *40*(5). https://doi.org/10.1007/s12664-021-01232-1

REBBECK, T. R. (2018). Prostate Cancer Genetics: Variation by Race, Ethnicity, and Geography. *Semin Radiat Oncol.*, *27*(1). https://doi.org/10.1016/j.semradonc.2016.08.002

REPS, J. M., SCHUEMIE, M. J., SUCHARD, M. A., RYAN, P. B., RIJNBEEK, P. R. (2018). Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *Journal of the American Medical Informatics Association, 25*(8), 969–975. https://doi.org/10.1093/jamia/ocy032

ROOBOL, M. J., VUGT, H. A., LOEB, S., ZHU, X., BUL, M., BANGMA, C. H., LEENDERS, A. G. L. J. H., STEYERBERG, E. W., SCHRÖDER, F. H. (2012). Prediction of Prostate Cancer Risk: The Role of Prostate Volume and Digital Rectal Examination in the ERSPC Risk Calculators. *European Urology*, *61*(3). https://doi.org/10.1016/j.eururo.2011.11.012

SANDHU, S., MOORE, C. M., CHIONG, E., BELTRAN, H., BRISTOW, R. G., WILLIAMS, S. G. (2021). Prostate cancer. *Lancet*, *398*. https://doi.org/10.1016/S0140-6736(21)00950-8

SEIBERT, T. M., GARRAWAY, I. P., PLYM, A., ..., MORGAN, T. M. (2023). Genetic Risk Prediction for Prostate Cancer: Implications for Early Detection and Prevention. *European Urology*, *83*(3). https://doi.org/10.1016/j.eururo.2022.12.021

SCHUMACHER, F. R., AL OLAMA, A. A., BERNDT, S. I., ..., ELLIPSE. (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nature Genetics*, *50*. https://doi.org/10.1038/s41588-018-0142-8

SHI, Z., PLATZ, E. A., WEI, J., ..., XU, J. (2021). Performance of Three Inherited Risk Measures for Predicting Prostate Cancer Incidence and Mortality: A Population-based Prospective Analysis. *Eur Urol*., *79*(3).

ST-HILAIRE, S., MANNEL, S., COMMENDADOR, A., MANDAL, R., DERRYBERRY, D. (2010). Correlations between meteorological parameters and prostate cancer. *Int J Health Geogr*., *9*(19). https://doi.org/10.1186/1476-072X-9-19

STEYERBERG, E. W., ROOBOL, M. J., KATTAN, M. W., KWAST, T. H., KONING, H. J., SCHRÖDER. (2007). Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. *J Urol*., *177*(1). https://doi.org/10.1016/j.juro.2006.08.068

STEYERBERG, E. W. (2019). Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating, Second Edition. *Springer*.

STRÖM, P., NORDSTRÖM, T., ALY, M., EGEVAD, L., GRÖNBERG, H., EKLUND, M. (2018). The Stockholm-3 Model for Prostate Cancer Detection: Algorithm Update, Biomarker Contribution, and Reflex Test Potential. *European Urology*, *74*(2).

TASA, T., PUUSTUSMAA, M., TÕNISSON, N., KOLK, B., PADRIK, P. (2020). Precision Prostate Cancer Screening with a Polygenic Risk Score. *medRxiv*. https://doi.org/10.1101/2020.08.23.20180570

TERVISE ARENGU INSTITUUT. (2023). Vähk Eestis: haigestumus 2020, elulemus 2016–2020 ja hematoloogilised kasvajad 2011–2020. *Tervise Arengu Instituut*.

THOMPSON, I. M., ANKERST, D. P., CHI, C., ..., COLTMAN, C. A. JR. (2006). Assessing Prostate Cancer Risk: Results from the Prostate Cancer Prevention Trial. *Journal of the National Cancer Institute*, *98*(8). https://doi.org/10.1093/jnci/djj131

VARMA, A., MAHARJAN, J., GARIKIPATI, A., HURTADO, M., SHOKOUHI, S., MAO, Q. (2023). Early prediction of prostate cancer risk in younger men using polygenic risk scores and electronic health records. *Cancer Med*, *12*(1).

VESKIMÄE, P., ŽARKOVSKI, M., KIVI, M., KIIVET, R. (2020). Eesnäärmevähi varane avastamine. TTH49. Tartu Ülikooli peremeditsiini ja rahvatervishoiu instituut.

VICKERS, A., O'BRIEN, F., MONTORSI, F., GALVIN, D., BRATT, O., CARLSSON, S., CATTO, J. W. F., KRILAVICIUTE, A., PHILBIN, M., ALBERS, P. (2023). Current policies on early detection of prostate cancer create overdiagnosis and inequity with minimal benefit. *BMJ*, *381*. https://doi.org/10.1136/bmj-2022-071082

VIETRI, M. T., D'ELIA, G., CALIENDO, G., RESSE, M., CASAMASSIMI, A., PASSARIELLO, L., ALBANESE, L., CIOFFI, M., MOLINARI, A. M. (2021). Hereditary Prostate Cancer: Genes Related, Target Therapy and Prevention. *Int J Mol Sci*, *22*(7). https://doi.org/10.3390/ijms22073753

VOS, I. I., MEERTENS, A., HOGENHOUT, R., REMMERS, S., ROOBOL, M. J. (2023). A Detailed Evaluation of the Effect of Prostate-specific Antigen–based Screening on Morbidity and

Mortality of Prostate Cancer: 21-year Follow-up Results of the Rotterdam Section of the European Randomised Study of Screening for Prostate Cancer. *European Urology*, *84*(4), 426–434. https://doi.org/10.1016/j.eururo.2023.03.016

WHO. (2004). ICD-10: international statistical classification of diseases and related health problems, tenth revision, 2nd ed. World Health Organization. https://iris.who.int/handle/10665/42980

XU, J., SUN, J., KADER, A. K., ..., GRÖNBERG, H. (2009). Estimation of Absolute Risk for Prostate Cancer using Genetic Markers and Family History. *Prostate*, *69*(14).

YANG, X., KAR, S., ANTONIOU, A. C., PHAROAH, P. D. P. (2023). Polygenic scores in cancer. *Nature Reviews Cancer*, *23*. https://doi.org/10.1038/s41568-023-00599-x

ZHANG, Y. D., HURSON, A. N., ZHANG, H., ..., GARCIA-CLOSAS, M. (2020). Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. *Nature Communications*, *11*. https://doi.org/10.1038/s41467-020-16483-3

ZHENG, S. L., SUN, J., WIKUND, F., ..., GRÖNBERG, H. (2008). Cumulative Association of Five Genetic Variants with Prostate Cancer. *N Engl J Med*., *358*(9). https://doi.org/10.1056/NEJMoa075819

Kasutatud veebiaadressid

CANCER RESEARCH UK. (2022). Prostate cancer. https://www.cancerresearchuk.org/about-cancer/prostate-cancer (vaadatud mai 2024)

EESTI TEADUSAGENTUUR. (2022). RITA MAITT lõpparuanne. https://www.etag.ee/wp-content/uploads/2022/05/RITA_MAITT_LOPPARUANNE_FINAL.pdf (vaadatud mai 2024)

EUROOPA VÄHITEABESÜSTEEM / EUROPEAN CANCER INFORMATION SYSTEM. (2023). https://ecis.jrc.ec.europa.eu/ (vaadatud mai 2024)

IQWIG. (2022). In brief: How does the prostate work? NCBI. https://www.ncbi.nlm.nih.gov/books/NBK279291/ (vaadatud mai 2024)

NCI. (2024). Genetics of Prostate Cancer – Health Professional Version. *National Cancer Institute*. https://www.cancer.gov/types/prostate/hp/prostate-genetics-pdq (vaadatud mai 2024)

NHS. (2022). Constitutional (germline) vs somatic (tumour) variants. NHS England. https://www.genomicseducation.hee.nhs.uk/genotes/knowledge-hub/constitutional-germline-vs-somatic-tumour-variants/ (vaadatud mai 2024)

OHDSI. (2019, uuendatud 2021). The Book of OHDSI. GitHub. https://ohdsi.github.io/TheBookOfOhdsi/ (vaadatud mai 2024).

OHDSI. (2024a). ATLAS. GitHub. https://github.com/OHDSI/Atlas

OHDSI. (2024b). ATHENA. GitHub. https://github.com/OHDSI/Athena

PADRIK, P. (2020). Peeter Padrik: mida saame teha enne vähktõve ravi. Postimees. https://www.postimees.ee/6875624/peeter-padrik-mida-saame-teha-enne-vahktove-ravi (vaadatud mai 2024)

PUNAB, M. (2006). Mehe teine süda. *Eesti Vähiliit*, Tallinn. https://www.digar.ee/arhiiv/nlib-digar:6981 (vaadatud mai 2024)

PUNAB, M. (2023a). Margus Punab selgitab, miks ei tohi kusemishäirete raviga venitada. Tervis Pluss. https://tervispluss.delfi.ee/artikkel/120232725/meeste-kusemishaired-pole-seotud-ainult-eesnaarmehaigustega-margus-punab-selgitab-miks-ei-tohi-kusemishairete-raviga-venitada (vaadatud mai 2024)

PUNAB, M. (2023b). Eesnäärme suurenemist ja -vähki leitakse üha noorematel meestel. Doktor Margus Punab selgitab, millal on viimane aeg kontrolli tulla. Tervis Pluss. https://tervispluss.delfi.ee/artikkel/120226602/eesnaarme-suurenemist-ja-vahki-leitakse-uha-noorematel-meestel-doktor-margus-punab-selgitab-millal-on-viimane-aeg-kontrolli-tulla (vaadatud mai 2024)

RIJNBEEK, P, REPS., J., RYAN., P., WILLIAMS, R. (2019). OHDSI Tutorial: Patient-level predictive modelling in observational healthcare data. Ettekanne. https://www.ohdsi-europe.org/images/symposium-2019/tutorials/PLP-Tutorial-All-2019-v1.pdf (vaadatud mai 2024)

SCHUEMIE, M., SUCHARD, M., RYAN, P., REPS, J., SENA, A., INBERG, G. (2024). FeatureExtraction: Generating Features for a Cohort. Versioon 3.5.2. GitHub. https://ohdsi.github.io/FeatureExtraction/index.html (vaadatud mai 2024)

Lihtlitsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Telver Objärtel

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Eesnäärmevähi riskimudeli arendamine Eesti terviseandmete põhjal“, mille vastutav juhendaja on Mart Kals, 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 Commonsi litsentsiga CC BY NC ND 4.0, 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.

Telver Objärtel
30.05.2024