

TARTU ÜLIKOOL

LOODUS- JA TÄPPISTEADUSTE VALDKOND

MATEMAATIKA JA STATISTIKA INSTITUUT

Sten Raak

**Puukborrelioos: geneetilised riskitegurid ja  
nakkusega seotud terviseriskid TÜ Eesti  
geenivaramu andmete põhjal**

Matemaatiline statistika

Bakalaureusetöö (9 EAP)

Juhendaja: PhD Erik Abner

Kaasjuhendaja: PhD Krista Fischer

TARTU 2023

**PUUKBORRELIOOS: GENEETILISED RISKITEGURID JA  
NAKKUSEGA SEOTUD TERVISERISKID TÕ EESTI  
GEENIVARAMU ANDMETE PÕHJAL**

Bakalaureusetöö

Sten Raak

**Lühikokkuvõte**

Bakalaureusetöö eesmärk on uurida puukborreliosisiga kaasnevaid terviseriske ning nakkusest tulenevaid haigestumISRISKE. Soovitakse leida põhjapanevaid tulemusi, mida Eesti meditsiinisüsteem saaks kasutada raskemate haiguste ennetuseks. Töös kasutatakse peamiselt hii-ruut testi ning logistilist regressiooni seose tugevuse hindamiseks. Lisaks on rakendatud eraldi PheW-ASi, mille järel on kasutatud Mendeli randomiseerimist põhjusliku seose tugevuse jaoks.

Esmalt antakse töös ülevaade puukborreliosisist ning sellega tulenevatest terviseprobleemidest. Seejärel kirjeldatakse seoseuuringute tüüpe GWAS ja PheWAS, miks neid kasutatakse ning nende erinevusi. Lõpuks tehakse kokkuvõte töös kasutatavatest meetoditest, näiteks hii-ruut test, logistiline regressioon, mitmese testimise probleem ning Mendeli randomiseerimine. Nende meetodite põhjal teostatakse analüüs puukborreliosisi ning puukborreliosisi soodustavale geneetilise riskiteguri mõjule teistele haigustele, kasutades Tartu Ülikooli Eesti geenivaramu andmeid.

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

**Märksõnad:** geneetilised assotsiatsiooniuringud, regressioon-analüüs, Mendeli randomiseerimine, mitmene testimine, puukborreliosis

**LYME DISEASE: GENETIC RISK FACTORS AND HEALTH RISKS  
ASSOCIATED WITH INFECTION BASED ON ESTONIAN  
GENOME RESPOSITORY OF THE UNIVERSITY OF TARTU  
DATA**

Bachelor thesis

Sten Raak

**Abstract**

The aim of the thesis is to investigate the health risks associated with Lyme disease and whether the infection increases the risk of other diseases. Another goal is to find fundamental results that can be used by general practitioners for prevention of serious diseases. The thesis mainly uses the chi-square test and logistic regression to assess the strength of the association. In addition, PheWAS is used separately for analysis. Finally, Mendelian randomization has then been applied to analyze the strength of the causal association.

Preliminarily, the work provides an overview of Lyme disease and its symptoms. Then there are described the types of association studies GWAS and PheWAS and the differences between them, and why they are used. Finally, a summary of all the methods used, such as the chi-square test, logistic regression, the multiple testing problem and Mendelian randomization, is given. Based on these methods, an analysis of the effect of genetic risk factor for Lyme disease and the effect of Lyme disease on other diseases will be carried out using data from the Estonian Genome Repository of the University of Tartu.

**CERCS research specialisation:** P160 Statistics, operation research, programming, actuarial mathematics

**Key Words:** Genetic association studies, regression-analysis, Mendelian randomization, multiple testing, Lyme disease

# Sisukord

<b>Sissejuhatus</b>	<b>5</b>
<b>1 Puukborrelioos</b>	<b>6</b>
1.1 Puukborrelioosi kirjeldus . . . . .	6
1.2 Diagnoosimine . . . . .	7
1.3 Ravijärgne Lyme'i tõve sündroom . . . . .	8
<b>2 Seoseuuringud ja nende erinevused</b>	<b>10</b>
2.1 Ülegenoomne seoseuuring . . . . .	10
2.2 Fenotüübi-põhine seoseuuring . . . . .	11
<b>3 Metoodika</b>	<b>13</b>
3.1 Hii-ruut test . . . . .	13
3.2 Logistiline regressioon . . . . .	14
3.3 Logistilise regressiooni tõlgendamine . . . . .	14
3.4 Mitmese testimise probleem . . . . .	15
3.5 Mendeli randomiseerimine . . . . .	16
<b>4 Geenivaramu andmete analüüs</b>	<b>19</b>
4.1 P53L ja puukborrelioosi vaheline seos . . . . .	19
4.2 PheWAS-analüüs geenivaramu andmetel . . . . .	21
4.3 Puukborrelioosiga kaasnevad diagnoosid . . . . .	25
4.4 Mendeli Randomiseerimine geenivaramu andmetel . . . . .	27
<b>Kokkuvõte</b>	<b>30</b>

Kasutatud allikad	32
Lisa 1. Puukborrelioosiga kaasnevate haiguste tabel	35

## Sissejuhatus

Puukborrelioos on puukide poolt levitatav haigus ning Eesti on selle endemiline piirkond. Puukborrelioosi diagnooside arv on 21. sajandil tugevalt kasvanud ning see nakkushaigus võib tekitada püsivaid kahjustusi, mistõttu on oluline uurida puukborrelioosi mõju teistele kaasuvatele haigustele. Varasemalt on leitud, et borrelioosiga kaasneb hulk sümptomeid, mille püsivuse kohta puuduvad kindlad teadmised.

Bakalaureusetöö põhieesmärk on uurida Tartu Ülikooli Eesti geenivaramu andmete põhjal, kas puukborrelioosi või puukborrelioosi nakatumist soodustava inimeeni ja teiste haiguste diagnooside vahel on olemas seos. Teine eesmärk on uurida, kas käesolev seos on põhjuslik või mitte.

Töös kasutatakse seoste uurimiseks peamiselt hii-ruut testi ning logistilist regressiooni. Lisaks kasutatakse fenotüübi-põhise seoseuuringu (PheWAS) meetodikat, mis põhineb logistilisel regressioonil, et samal ajal uurida ühe geenivariandi seoseid mitmete erinevate haigustega. Viimasena on rakendatud ka Mendeli randomiseerimist, et uurida seoste põhjuslikkust ja vähendada segavate faktorite mõju.

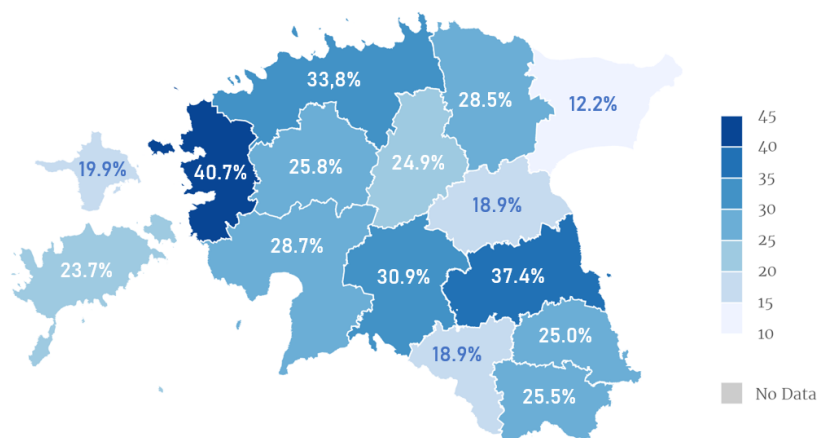
Esimeses peatükis antakse ülevaade puukborrelioosist, et lugejale oleksid arusaadavad haiguse olemus ja sümptomid. Teises peatükis on selgitatud, mis põhimõttel töötavad kaks erinevat seoseuuringut (GWAS ja PheWAS) ning kuidas neid kasutatakse. Kolmandas peatükis selgitatakse kõiki statistilisi meetodeid, mida töös analüüsimisel on kasutatud. Neljas peatükk algab SCGB1D2:P53L/rs2232950 geenivariandi ja puukborrelioosi seose analüüsiga ja seejärel kirjeldatakse kõikide teiste analüüside protsesse ning tulemusi.

Andmetöötlusteks ning visualiseerimiseks kasutati statistikatarkvara R 4.2.2 (R Core Team, 2022). Töö on vormistatud LaTeX-s läbi Overleafi veebikeskkonna.

# 1 Puukborrelioos

## 1.1 Puukborrelioosi kirjeldus

Puukborrelioos (PB) ehk Lyme'i tõbi on nakkushaigus, mida põhjustavad *Borrelia* perekonda kuuluvad bakterid (peamiselt *Borrelia burgdorferi*) ning mis levivad puukide kaudu (Joonis 1). Enamasti on nakkus lühiajaline, kuid see võib areneda raskemasse faasi, mis nõuab intensiivset antibiootikumiravi ja võib viia krooniliste haiguste tekkeni. Viimastel aastakümnetel on borrelioosi nakatunute arv järjest kasvanud, seda osaliselt tingitud tänapäeva meditsiini võimekusest puukborrelioosi usaldusväärsemalt tuvastada võrreldes varasemaga. (Strausz *et al.*, 2022)



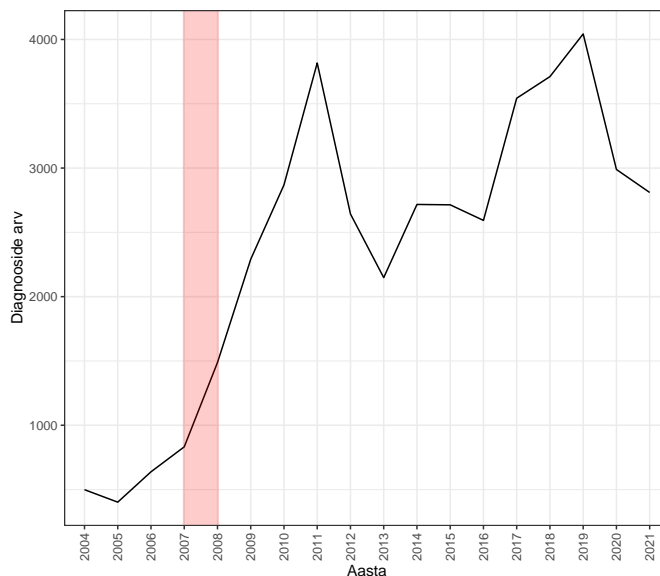
Joonis 1: *Borrelia* bakterite osakaal puukides maakondade kaupa aastal 2020. Andmed: Tervise Arengu Instituut

Puukborrelioosi kannavad edasi puugid, kelle organismi on bakter jõudnud nakatunud loomade vere kaudu. Puugid ise ei nakatu, aga kannavad seda edasi teistele loomadele ja ka inimestele. Selle nakkushaigusega kõige sagedamini kaasnevad sümptomid on peavalu, palavik, nõrkustunne, lihas- ja liigesevalud ning kaalulan-  
gus. Lisaks on iseloomulik omapärase välimusega nahalööve (*erythema migrans*), mis väljendub vähemalt 5 cm läbimõõduga põletikulise punase ringina. (Tervise-  
amet, kuupäev puudub)

## 1.2 Diagnoosimine

Puukborrelioosi diagnoosimine sõltub peamiselt sümptomite ja puugirünnaku tuvastamisest. Kuigi laboratoorsed uuringud aitavad kinnitada tüüpilist kliinilist leidu, ei põhine borrelioosi diagnoos kunagi ainult laboratoorsetel parameetritel või ebamääraste sümptomite esinemisel. Puukborrelioosi kahtluse korral uuritakse esmalt IgM ja IgG tüüpi antikehade olemasolu. Positiivse tulemuse korral rakendatakse teisena immunoblottimist. (Aguero-Rosenfeld ja Wormser, 2015)

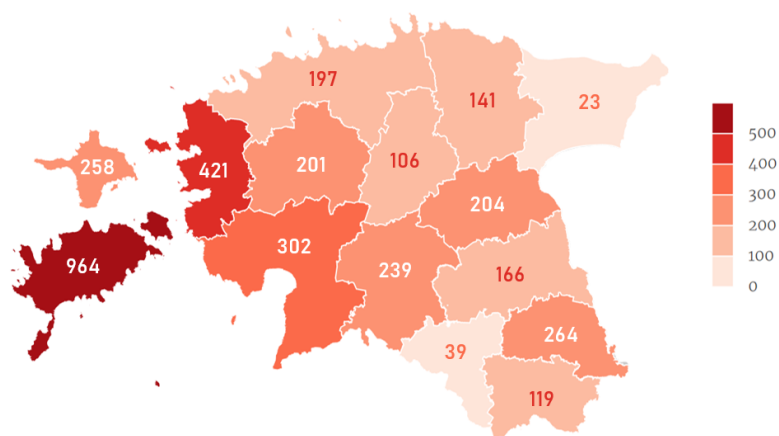
Seroloogilised testid on Eestis igapäevases kliinilises praktikas kergesti kättesaadavad. Need testid võeti kasutusele 2007-2008 vahemikus, mille tagajärjel tõenäoliselt kaasnes suurenenud diagnooside arv (Joonis 2). Samas tuleb puukborreliooside diagnooside arvu hindamisel arvestada segavate faktoritega, mis võivad tulemusi mõjutada.



Joonis 2: Diagnooside arv geenidoonorite seas vahemikus 2004-2021

Eestis kasutatakse semikvantitatiivseid ELISA-teste, mille tulemusena saadakse antikehade tiiter. Kahjuks saab neid teste usaldusväärset hinnata ainult positiivse või negatiivse tulemuse põhjal. Nõrgalt positiivsed tulemused on ebatäpsed, seega

teste tuleks mõne nädala pärast korrata. Tulemuste tõlgendamisel ELISA-ga tuleb arvestada, et paljudel inimestel on borrelioosivastased antikehad veres ka siis, kui nad pole endale teadaolevalt hiljuti puukborrelioosi põdenud. Eriti tihti juhtub seda piirkondades, kus on palju puuke, näiteks Lääne-Eestis (Joonis 3). Pärast borrelioosi põdemist võivad antikehad püsida inimese kehas aastaid, mis omakorda võib põhjustada valepositiivsete diagnooside arvu suurenemist. (Prükk, Maimets ja Lutsar, 2012)



Joonis 3: Puukborrelioosi juhud 100 000 elaniku kohta maakondade kaupa aastal 2021. Andmed: Terviseamet

### 1.3 Ravijärgne Lyme'i tõve sündroom

Enamasti ravitakse puukborrelioos 3-4 nädalaga, kasutades suukaudset antibiootikumikuuri. Siiski võib patsientidel aeg-ajalt tekkida valu, väsimus ning tähelepanuraskused, mis võivad kesta üle 6 kuu pärast esialgse ravi lõpetamist (Ursinus *et al.*, 2021). Seda seisundit nimetatakse ravijärgseks Lyme'i tõve sündroomiks (inglise keeles *Post Treatment Lyme Disease Syndrome*, edaspidi PTLDS). (Ścieszka, Dąbek ja Cieślík, 2015)

Täpselt ei ole teada, miks osad inimesed kogevad PTLDSi. Osa eksperte arvavad, et bakterid vallandavad autoimmuunse reaktsiooni, mis põhjustavad ka borrelioosi

sümptomeid peale selle ravi. Teised eksperdid arvavad, et sümptomid tulenevad latentsest püsivast infektsioonist. Lisaks on ka koolkond, kus usutakse, et PTLDSi haigusnähud on põhjustatud täiesti teistsugusest haigustest või probleemidest, mis ei ole kuidagi puukborrelioosiga seotud. (CDC ja Prevention, 2022)

Kuigi antibiootikumid aitavad puukborrelioosi ravida, ei soovitata neid pikaajaliselt tarvitada, sest need võivad tekitada tõsiseid tüsistusi. Kuna PTLDSi põhjused on veel tuvastamata ning ametlikud ravimeetodid puuduvad, siis on ravi sümptomaatiline. Enamasti peaksid kõik sündroomiga patsiendid tervenema ilma antibiootikumideta. (CDC ja Prevention, 2022)

## 2 Seoseuuringud ja nende erinevused

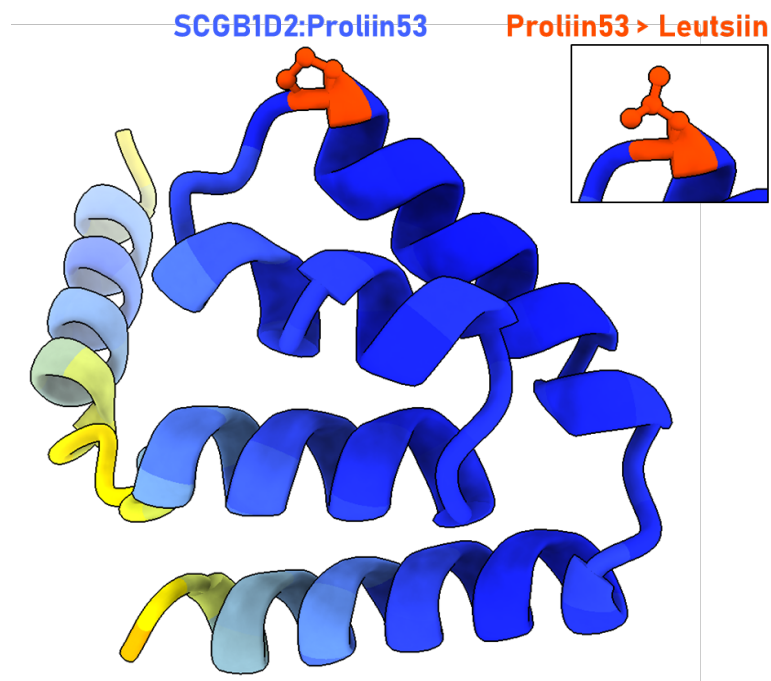
### 2.1 Ülegenoomne seoseuuring

Ülegenoomne seoseuuring (edaspidi GWAS, *genome-wide association study*) on meetod, mida kasutatakse konkreetsete haiguste või muude tunnuste riskidega seotud geenivariantide tuvastamiseks. Selle meetodiga uuritakse suure hulga inimeste genome selleks, et leida geenivariante, mis esinevad sagedamini teatud haigustega inimestel võrreldes nendega, kellel antud geenivarianti ei esine. (NIH, 2023)

Esimene GWAS avaldati aastal 2005, mille järgselt hakati meetodit laiahaardeliselt üle maailma kasutama (Ku *et al.*, 2010). Ülegenoomsetel seoseuuringutel on kasvanud aastatega valimi suurus - algselt rakendati meetodit mõne tuhande inimese peal, nüüd juba miljonite. (Uffelmann *et al.*, 2021)

Sageli viiakse GWAS-uuringud läbi tuhandete uuringus osalejate peal, mis hõlmavad geneetilist varieeruvust ühe nukleotiidi polümorfismide (SNP) kujul kogu inimgenoomis. Iga SNP-d, mille arv võib ulatuda miljonitesse, hinnatakse statistiliselt, et tuvastada seosed täpselt määratletud uuritava fenotüübiga. GWASi abil saadud tulemuste probleemiks on reaalse tagajärgede ja mõjude uurimine (Huang, 2015). GWASi andmeid tuleb analüüsida, võttes arvesse valepositiivsete tulemuste liiga suurt osakaalu arvukate statistiliste testide puhul. (NIH, 2023)

Soomes teostati GWAS puukborrelioosile, kus 5 248 inimesel oli esinenud puukborrelioos ja 333 251 inimest olid kontrollgrupis. Olulisteks tulemusteks saadi kaks geenivarianti: HLA-DQB1\*06:02 ning SCGB1D2:P53L (edaspidi P53L geenivariant) (Joonis 4).



Joonis 4: SCGB1D2 geenilt tuleneva valgu struktuur, kus punasega on selle muudatus P53L välja toodud. Allikas: AlphaFold (UniProt ID:O95969)

## 2.2 Fenotüübi-põhine seoseuring

Ülegenoomsete seoseuringute tulemusena on selgunud, et mõni geenivariant on seotud mitme erineva haiguse või tunnusega. Kui uuringuteks on võimalik kasutada suuremahulist fenotüübi-andmebaasi, siis on vahel osutunud otstarbekaks ka analüüsida ühe geenivariandi (SNP) seost suure arvu erinevate haigustega. (vastupidiselt GWASile, kus testitakse ühe fenotüübitunnuse seost paljude genotüüpidega). Seda nimetatakse fenotüübi-põhiseks seoseskaneeringuks (edaspidi PheWAS). (Denny *et al.*, 2010)

PheWASi saab rakendada GWASi tulemuste põhjal. On olukordi, kus esialgselt on kasutatud GWASi, et leida fenotüübitunnusele seost genotüüpidega, ning seejärel rakendatakse sellele genotüübile PheWASi, et uurida seoseid teiste haigustega. (Schlauch *et al.*, 2018)

Võimalus läbi viia PheWAS võib olla viis uute geneetiliste seoste avastamiseks ning haiguste mehhanismide tundmaõppimiseks. Eelkõige on PheWASi abil võimalik kindlaks määrata, millised haigused seostuvad mõne konkreetse SNP-iga. (Denny *et al.*, 2010)

## 3 Metoodika

### 3.1 Hii-ruut test

Hii-ruut testi kasutatakse kahe mitteamulise tunnuse vahelise seose kindlaksteegemiseks. See on levinud bioloogia, sotsiaalteaduste ja psühholoogia valdkonnas, et määrata, kas vaadeldud andmed sobivad teoreetilise jaotusega.

Test põhineb eeldusel, et kui kaks tunnust on sõltumatud, siis iga nende tunnuste väärtuste poolt määratud kategooria sagedus vaadeldud andmetes peaks olema ligikaudu võrdne oodatud sagedusega.

Kui kaks tunnust  $X$  ja  $Y$  on sõltumatud, siis iga võimaliku väärtuspaari ( $X = x, Y = y$ ) suhteline esinemissagedus võiks olla ligikaudu võrdne oodatava sagedusega  $P(X = x, Y = y) = P(X = x)P(Y = y)$ . Oodatud sagedus arvutatakse seetõttu kahe muutuja marginaalsageduste korrutisena.

Hii-ruut statistik leitakse valemiga:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i},$$

kus  $O_i$  on vaadeldud sagedus (*observed*) ning

$$E_i = \frac{n_{i.} \cdot n_{.j}}{n_{..}}$$

on oodatud sagedus (*expected*), kus  $n_{i.}$  ja  $n_{.j}$  tähistavad marginaalsagedusi kahe mõõtmelise sagedustabeli reas  $i$  ja veerus  $j$  ning  $n_{..}$  on vaatluste koguarv.

Nullhüpoteesi kehtimisel on see teststatistik  $\chi^2$ -jaotusega vabadusastmete arvuga  $(r - 1) \cdot (c - 1)$ , kus  $r$  on ridade arv ja  $c$  on veergude arv sagedustabelis. Seega kui statistiku väärtus jääb suuremaks selle jaotuse  $\alpha$ -täiendkvantiilist, saab olulisuse nivool  $\alpha$  otsustada, et vaadeldavad tunnused ei ole sõltumatud.

## 3.2 Logistiline regressioon

Uurides P53L geeni mõju haigestumisele, vaadeldakse haigestumist kui binaarset tunnust, mis uuritaval inimesel on 1, kui tal on see haigus esinenud, ja 0, kui seda ei ole esinenud. Haigestumise tõenäosuse või šansi hindamiseks sõltuvalt genotüübist, võttes lisaks arvesse ka teiste tunnuste mõju, nagu sugu ja vanus, saab kasutada logistilise regressiooni mudelit.

Olgu  $p_i$  haigestumise tõenäosus indiviidil  $i$ . Sellisel juhul saab kasutada uuritava sündmuse tõenäosuse kujutamiseks logit-seosefunktsiooni:

$$\text{logit}(p_i) = \ln \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki},$$

kus  $(\beta_0, \dots, \beta_k)$  on mudeli parameetrid (kordajad) ning  $(x_{1i}, \dots, x_{ki})$  on argument-tunnused vastavalt mitme erineva tunnuse järgi sündmuse toimumise tõenäosust üritatakse hinnata. Mudeli parameetreid hinnatakse suurima tõepära meetodil.

## 3.3 Logistilise regressiooni tõlgendamine

Logit seosest saab avaldada sündmuse esinemise tõenäosuse ehk prognoosi tõenäosuse:

$$p_i = \frac{e^{\text{logit}(p_i)}}{1 + e^{\text{logit}(p_i)}}.$$

Sündmuse šanss on sündmuse esinemise tõenäosuse ja mitteesinemise tõenäosuse suhe:

$$\Pi_i = \frac{p_i}{1 - p_i}.$$

Šansside suhe (*odds ratio*) on kahe isiku või objekti sündmuste šansside jagatis. Juhul, kui parameetri argument muutub  $c$  võrra, siis muutub šansside suhe  $e^{c\beta}$  korda.

$$OR = \frac{\Pi_i}{\Pi_j} = \frac{\frac{p_i}{1-p_i}}{\frac{p_j}{1-p_j}}.$$

### 3.4 Mitmese testimise probleem

Mitmese testimise probleem tekib, kui sama valimi põhjal testitakse korraga mitmeid hüpoteese.

Statistikas on nullhüpotees ( $H_0$ ) hüpotees, mida püüakse ümber lükata. Nullhüpotees tähendab enamasti erinevuste puudumist kahe rühma vahel (näiteks geenil ei ole mõju uuritava tunnusele), aga alternatiivseks hüpoteesiks ( $H_1$ ) määratakse, et leidub erinevus kahe rühma vahel (näiteks geenil on mõju uuritavale tunnusele).

Statistilise testi korral on neli võimalikku tulemust:

- test lükkab ümber nullhüpoteesi siis, kui kehtib alternatiivne hüpotees;
- test ei lükka ümber nullhüpoteesi, kui kehtib nullhüpotees;
- test lükkab ümber nullhüpoteesi siis, kui kehtib nullhüpotees (I liiki viga);
- test ei lükka ümber nullhüpoteesi, kui kehtib alternatiivne hüpotees (II liiki viga);

Näiteks kui testida valimis 20 erineva tunnuse seost haiguse esinemisega olulisuse nivool 0,05, kuid tegelikult pole mitte ükski neist tunnustest haigusega seotud (kehtib nullhüpotees), siis võib oodata, et keskmiselt üks tunnus 20-st annab siiski statistiliselt olulise seose - tegelikult on see leid valepositiivne. Sama põhimõtte järgi suureneb 20 000 geeni analüüsimisel oodatavate valepositiivsete tulemuste arv kuni 1 000-ni. Üheaegsete statistiliste testide arvu tõstmine suurendab ohtu, et valepositiivsed tulemused ületavad reaalsete avastuste arvu. Seetõttu on vajalik rakendada meetmeid, et vähendada valepositiivseid tulemusi. (Menyhart, Weltz ja Györfy, 2021)

Benjamini-Hochbergi kirjeldatud valeavastamise määr tagab õige tasakaalu statistiliselt oluliste mõjude avastamise ja valepositiivsete piirangute vahel. Praktikas põhineb protseduur p-väärtuste järjestamisel kasvavas järjekorras, pärast mida ar-

vutatakse iga individuaalse  $p$ -väärtuse Benjamini-Hochbergi kriitiline väärtus:

$$p_{B_i} = \frac{p \cdot i}{n},$$

kus  $i$  on testi järjekorranumber,  $n$  on testide arv kokku,  $p$  on olulisuse nivoo ning  $p_{B_i}$  on Benjamini-Hochbergi kriitiline väärtus, millest testi olulisuse tõenäosus peab väiksem olema, et alternatiivset hüpoteesi vastu võtta. Tulemuseks on testidel suurem võimsus, kui näiteks Bonferroni korrektsiooni kasutada. (Benjamini ja Hochberg, 1995)

### 3.5 Mendeli randomiseerimine

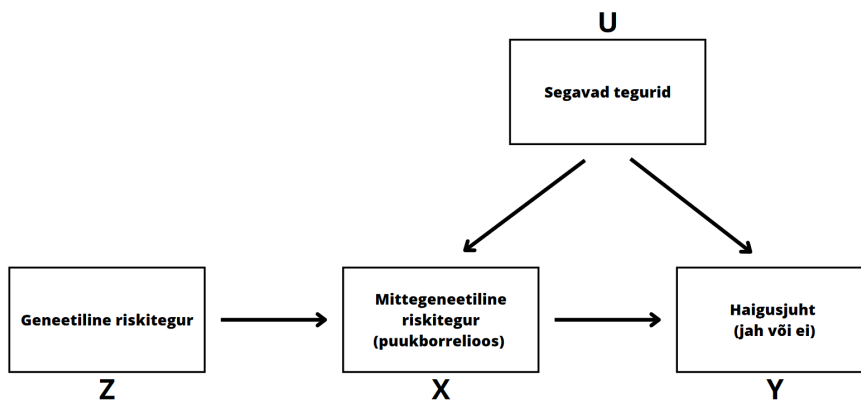
Mendeli randomiseerimine (MR) on riskifaktori ja uuritava tunnuse vahelise põhjusliku suhte uurimise meetod, kus kasutatakse geneetilisi andmeid instrumenttunnustena. See lähenemine tugineb Mendeli seadusele, kus geenid jaotatakse juhuslikult ega allu segavatele tunnustele või tagurpidisele põhjuslikkusele. (Davey Smith ja Hemani, 2014)

MR on saanud väga populaarseks hiljuti ning seda kasutatakse mitmetes valdkondades, näiteks geneetikas, meditsiinis või ka metabolismi uuringutes. Traditsioonilistest vaatlusuuringutest erineb MR selle poolest, et antud meetod võimaldab teha järeldusi seoste põhjuslikkuse kohta. MR võib anda teavet võimalike ravieesmärkide kohta ja suunata ravimeetodite väljatöötamist, et parendada tervisetulemusi. (Larsson, 2021)

MR-i läbiviimiseks peavad teadlased tuvastama geenid, mis on seotud uuritava riskifaktoriga. Kui neil geenidel tuvastatakse lisaks ka oluline seos uuritava tulemusega (nt haigusega), siis võib sealt järeldada riskifaktori põhjuslikku mõju tulemusele. Tulemuste kehtivuse tagamiseks on sellise lähenemisviisiga vaja suurt valimit ja rangeid statistilisi meetodeid. (Davey Smith ja Hemani, 2014)

Meetodi põhimõtte tugineb allolevale skeemile (Joonis 5), kus üritatakse leida mitte-

geneetilise riskiteguri põhjuslikku mõju. Meetodi kasutamisel peab kontrollima ka kolme eeldust. Esimesena peab eksisteerima seos geneetilise riskiteguri ja mittegeneetilise riskiteguri vahel. Teiseks peab geneetilisel riskiteguri puuduma põhjuslik mõju uuritavale haigusele, arvestades ka kõiki segavaid faktoreid. Kolmandaks ei tohi geneetilisel riskiteguri olla seost segavate faktoritega, mis võivad mittegeneetilist riskiteguri ja haigust mõjutada. (VanderWeele *et al.*, 2014)



Joonis 5: Mendeli Randomiseerimise skeem

MR parameetri avaldamiseks koostame kaks võrrandit (kasutades joonisel olevaid tähiseid):

$$E(X|Z) = \alpha_0 + \alpha_1 Z$$

$$E(Y|X) = \beta_0 + \beta_1 X + \gamma U, \quad \text{eeldusel, et } U \perp Z$$

Siin  $\alpha_1$  tähistab geeni mõju mittegeneetilisele riskitegurile,  $\beta_1$  mittegeneetilise riskiteguri mõju haigusele ning  $\gamma U$  segavate faktorite mõju. Nüüd on võimalik avaldada  $E(Y|Z)$  sellisena:

$$E(Y|Z) = \beta_0^* + \beta_1 E(X|Z) = \beta_0^* + \beta_1 \alpha_1 Z,$$

kasutades  $E(Y|Z)$  ja  $E(X|Z)$  parameetrite jagatist, saab  $\beta_1$  leida:

$$\beta_1 = \frac{\beta_1 \alpha_1 \cdot Z}{\alpha_1 \cdot Z} = \frac{\beta_1 \alpha_1}{\alpha_1},$$

mis hinnanguna avaldub:

$$\hat{\beta}_1 = \frac{\beta_1 \hat{\alpha}_1}{\hat{\alpha}_1}$$

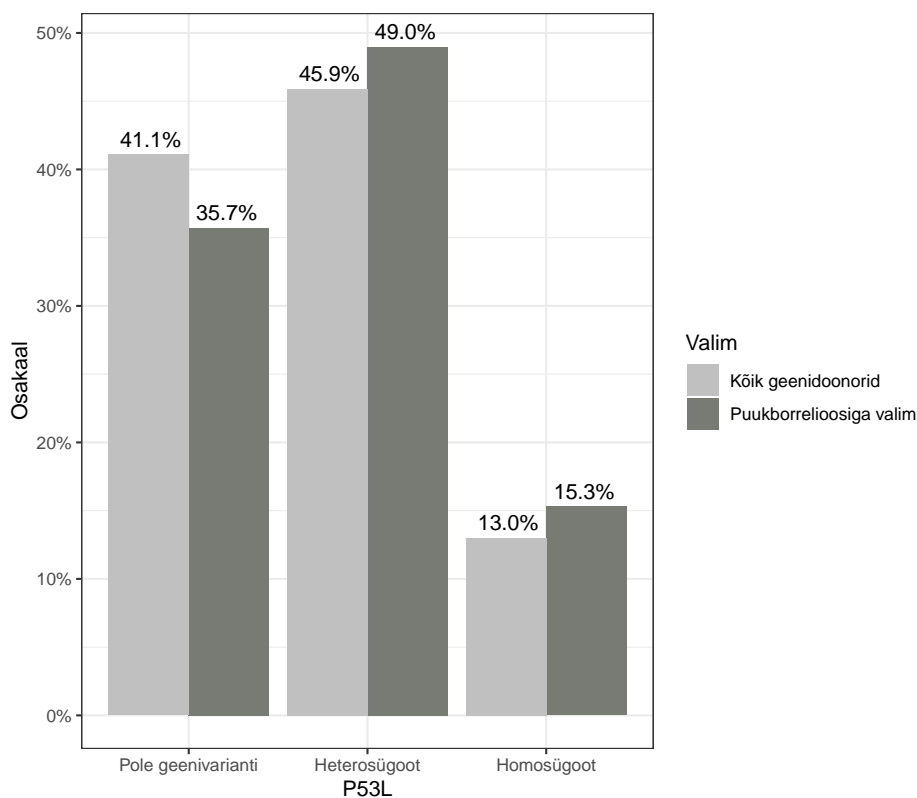
Üldiselt on MR efektiivne meetod, mis võib anda suurema tähendusega tulemusi kui lihtsalt seoseid uurides. Tihti soovitakse leida põhjuslikku seost kahe tunnuse vahel. Meetodi kasutamise teevad probleemseks teatud piirangud ning mitmed eeldused, mis peavad täidetud olema, et tulemuste põhjal saaks järeldusi teha.

## 4 Geenivaramu andmete analüüs

### 4.1 P53L ja puukborrelioosi vaheline seos

Geenivaramu andmestikus on 18 977 inimest, kes on saanud ajavahemikus 2004-2021 puukborrelioosi diagnoosi. Nendest 17 655-l on määratud P53L genotüüp. Inimesi, kellel on määratud P53L genotüüp, on kokku 202 909 (Geenivaramu andmeväljastuse versioon 2022v02).

Varasematest uuringutest on teada, et P53L geenivariant suurendab puukborrelioosi haigestumise riski (Strausz *et al.*, 2022). Seda on võimalik näha ka andmetes, kui näiteks võrrelda genotüübi jaotust puukborrelioosiga diagnoositute ning geenivaramu kõikide geenidoonorite seas (Joonis 6).



Joonis 6: P53L geenivariandi arvukuse osakaalude võrdlus

Jooniselt on näha, et populatsioonis on P53L geenivariandiga inimeste osakaal väiksem kui puukborrelioosiga grupis. Kahe tunnuse seose uurimiseks on andmed agregeeritud kokku sagedustabeliks (Tabel 1) ning sellele rakendatud hii-ruut testi.

Tabel 1: Genotüübi ja puukborrelioosi diagnooside sagedustabel

Riskialleelide arv	Ei ole PB diagnoositud	On diagnoositud PB
0	77 095	6 297
1	84 491	8 657
2	23 668	2 701

Hii-ruut testi rakendamine eelneva tabeli peal annab tulemuseks p-väärtuse, mis on väiksem kui  $2,2 \cdot 10^{-16}$ . Seega saab öelda, P53L ja PB nakatumise vahel leidub oluline seos. Selleks, et paremini aru saada, kuidas geenide arv mõjutab nakatumist, on välja arvatatud šansside suhted (Tabel 2).

Tabel 2: Šansside suhted võrreldes 0 riskialleeliga

Riskialleelide arv	Šansside suhe (95% UI)
1	1,25 (1,21; 1,30)
2	1,40 (1,33; 1,47)

Heterosügootsel P53L geenivariandi kandjal on 1,25 korda suurem šanss nakatuda puukborrelioosi kui ilma riskialleelita inimesel. Samas homosügootsel P53L geenivariandi kandjal on 1,4 korda suurem võimalus puukborrelioosi nakatuda.

Puukborrelioosivastaste antikehade pika poolestusaja ja sellest tulenevate seroloogiliste uuringute ebatäpsuse tõttu on viimastel aastatel borrelioosi ülediagnoosimise võimalus (Kobayashi *et al.*, 2022). Tihti lisatakse puukborrelioos kõrvaldiagnoosina, kui patsient kurdab puukborrelioosiga kaasneva sümptomi üle.

Nüüd uuritakse, kas eelnevalt mainitud geeni pärimine mõjutab borrelioosi peadiagnoosi saamist. Hii-ruut test näitab, et geenivariandi ja peadiagnoosi saamise vahel ei ole seost olemas ja p-väärtus on 0,3505.

Vastupidise seose uurimine annab infot, kas P53L puudumine võiks mõjutada kõrvaldiagnooside saamist. Selle jaoks on püstitatud küsimus: Kas P53L riskialleelide arv mõjutab kõrvaldiagnoosi saamist? Hii-ruut testiga tuleb p-väärtuseks 0,4325. Nullhüpoteesi ei lükata ümber ning jääb kehtima seisukoht, et riskialleelide arv ei mõjuta puukborrelioosi kõrvaldiagnoosi saamist.

## 4.2 PheWAS-analüüs geenivaramu andmetel

Esiialgu uuritakse, kas puukborrelioos mõjutab meie andmetes teadaolevaid borrelioosiga seonduvaid haigusi ehk kas puukborrelioosi põdemine suurendab teiste raskemate haiguste riski. Mõni haigus võib olla puukborrelioosiga seotud sellepärast, et sellel haigusel on puukborrelioosiga sarnased riskitegurid, nagu sage looduses viibimine.

Antud andmetel põhjal on võimalik teostada PheWAS analüüs. PheWAS analüüs sobib, sest olemas on uuritav sündmus (haigus), mõjutav tunnus (puukborrelioos või P53L genotüüp) ning kovariaadid (vanus, sugu, geneetilised peakomponendid).

Suure tõenäosusega võib olla segav faktor näiteks geneetiline sugulus. Tuleb arvestada sellega, et perekond võib olla sarnase eluviisiga ja perekond, kes käib metsas, saab suurema tõenäosusega puuke endale külge. Teine näide segavast faktorist on vanus. Vanemal inimesel on võimalik rohkem haigusi põdeda võrreldes nooremaga.

Uuritavaks tunnuseks on võetud kõik teised haigused peale puukborrelioosi. Tunnus, mille mõju tahetakse teada saada, on kas P53L geenivariandi või puukborrelioosi esinemine. Puukborrelioosi korral pole formaalselt küll tegemist PheWAS-ga, aga kuna analüüs on identne PheWAS-ga, siis kasutati sama tarkvara. Kovariaatideks on määratud geneetilised peakomponendid, mis kirjeldavad, kui geneetiliselt sarnased on geenidonorid. Nendele lisaks on kovariaatideks veel vanus ning sugu.

Rakendustarkvaras R on pakett nimega *PheWAS*, mille abil saab automaatselt rakendada PheWAS meetodikat. Paketis on mitmeid funktsioone, mis abistavad

andmete töötlemist ning mudeli rakendamist. Kõige raskem on viia kõik andmed sellisele kujule, mida oleks võimalik *phewas\_ext()* funktsioonile ette anda. See on peamine funktsioon, mida kasutatakse, ning väljastab vajaliku informatsiooni ühe-realise tabelina.

Andmestikus on informatsioon geenidonorite diagnooside kohta aastatel 2004-2021. PheWAS analüüsi jaoks kasutatakse ligikaudu 10 000 erineva diagnoosi kohta informatsiooni binaarsel kujul, et kas geenidonor on saanud vahemikus 2004-2021 selle diagnoosi või mitte (filtreeritud, et juhtumeid oleks vähemalt 100). Diagnooside nimetused on kodeeritud RHK-10 klassifikaatori järgi, näiteks puukborrelioos on A69.2 (Sotsiaalministeerium, kuupäev puudub).

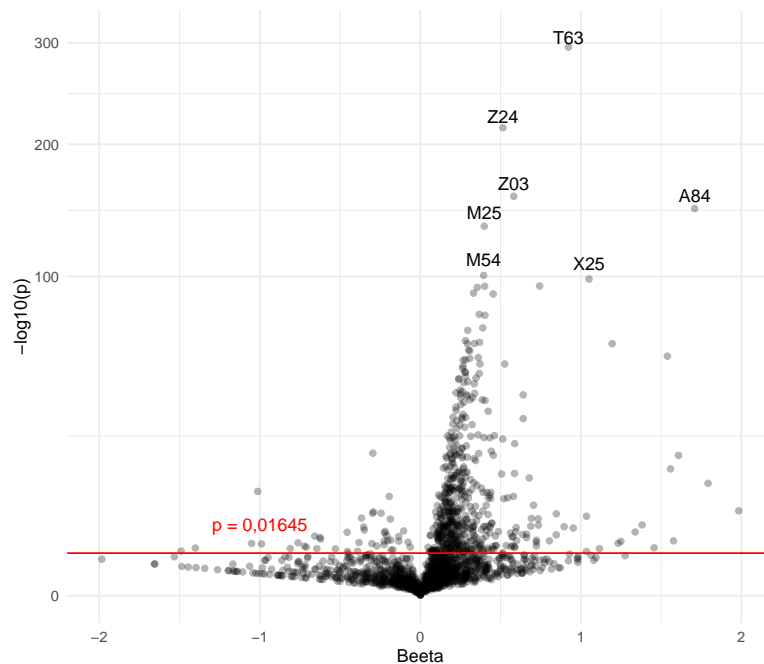
RHK-10 koodidest on elimineeritud kõik spetsiifilisemad koodid ehk alles on jäetud RHK-10 kategooriad, näiteks T63, mis tähistab mürgiste loomade toksilist toimet. Kui inimesele on määratud väga täpne diagnoos, siis andmestikus on märgitud, et tal on sel juhul esinenud ka üldisem variant sellest, et vähendada väga haruldasi ning spetsiifilisi diagnoose ning aitab üldpildist paremini aru saada. Alles jääb 1834 RHK-10 kategooriat, mida edaspidi PheWAS-s kasutatakse.

Selgema seose nägemiseks ning segavate faktorite eemaldamiseks on mudelisse lisatud kovariaatideks sugu, vanusevahemik ning 10 geneetilist peakomponenti. Lisaks on rakendatud Benjamin-Hochbergi p-väärtuse korrigeerimise. Allolevas tabelis on välja toodud esimesed viis olulist diagnoosi, mis on väikseima p-väärtuse järgi valitud (Tabel 3):

Tabel 3: PheWAS põhjal puukborrelioosi oluline mõju

RHK-10	Nimetus	Šansside suhe (95% UI)	P-väärtus
T63	Mürgiste loomade toksiline toime	2,52 (2,40; 2,64)	$4,83 \cdot 10^{-296}$
Z24	Immuniseerimise vajadus viiruse vastu	1,67 (1,62; 1,73)	$1,30 \cdot 10^{-215}$
Z03	Meditiiniline jälgimine/hindamine	1,79 (1,71; 1,87)	$2,91 \cdot 10^{-157}$
A84	Puukentsefaliidid (viirusentsefaliidid)	5,53 (4,86; 6,29)	$7,90 \cdot 10^{-148}$
M25	Muud liigeste haigusseisundid	1,49 (1,44; 1,54)	$1,30 \cdot 10^{-134}$

Analüüsi kõiki tulemusi vaadates ilmneb probleem: oluliseks jääb 561 haigust (Joonis 7), kus puukborrelioosi ja haiguse vahel on oluline seos. Joonisel märgib punase joone kõrgust kõige viimase olulise seose p-väärtus (0,01645). Siin võivad mängida rolli muud segavad tegurid, mis on näiteks ainult puukborrelioosiga inimestel ning mida mudelisse lisatud kovariaadid ei suuda elimineerida. Seega PheWASi tulemustega ei ole võimalik mõistlikke järeldusi teha, vaja oleks andmetele läheneda teistmoodi.



Joonis 7: PheWAS rakendamise tulemused A69.2 ja teiste diagnooside vahel (y-telg ruutjuurega transformeeritud)

Eelnevalt on teada, et P53L pärimine avaldab märkimisväärset mõju borrelioosi nakatumisele. Selle tõttu tasub ka vaadata, kas riskialleelide pärimisel on mõju teistesse haigustesse nakatumisele. Võimalik on uuesti rakendada PheWASi analüüsi, nüüd valides tunnuseks, mille mõju on vaja uurida, P53L genotüübi arvu. Paketis teisendatakse kõik suuremad arvud üheks. Selleks, et saada teada 2 alleeli mõju, peab šansside suhte ruutu võtma.

Kasutades PheWASi paketi funktsiooni on tulemuseks (Tabel 4) neli diagnoosi, millele P53L mõju avaldab ja mis jäävad oluliseks peale Benjamin-Hochbergi korrigeerimise. Kõik neli diagnoosi on seotud putukate või lüliljalgsetega kokku puutumise, mis võib tuleneda puugihammustustest ja seega olla seotud ka borrelioosiga.

Tabel 4: PheWAS põhjal P53L oluline mõju

RHK-10	Nimetus	Šansside suhe (95% UI)	P-väärtus
A69	Muud spiroheetnakkused	1,19 (1,17; 1,22)	$6,46 \cdot 10^{-53}$
T63	Mürgiste loomade toksiline toime	1,11 (1,08; 1,14)	$1,65 \cdot 10^{-13}$
W57	Mittemürgiste putukate hammustus/nõel	1,11 (1,07; 1,15)	$2,56 \cdot 10^{-08}$
X25	Kokkupuude mürgiste lüljalgsetega	1,14 (1,07; 1,20)	$1,64 \cdot 10^{-05}$

Kuigi P53L avaldas olulist mõju ainult neljale haigusele, ei ole kahjuks võimalik eelnevatest analüüsides mõjuvaid järeldusi teha. P53L puhul on kõik neli haigust seotud putukate hammustustega või nõelamistega, aga puukborreliosisil jäid olulisteks üle 500 haiguse. Oluliseks jäid diagnoosid, mis ei tundu puukborreliosisi või sellest põhjustatud tüsistustega seotud olevat. Üheks näiteks on õlakahjustused, mida esineb borreliosisi põdenutel 1,36 korda rohkem kui teistel.

### 4.3 Puukborreliosisiga kaasnevad diagnoosid

Järgnevalt on eesmärk teada saada, millised on põhilised puukborreliosisiga kaasnevad diagnoosid. Seda saab teha, kasutades täpsemaid andmeid arstivisiitide kohta (Tervisekassa). Teada on kõik diagnoosid, mis on pandud arstivisiidil ning tegu on kõrval- või peadiagnoosiga. Olemas on ka arstivisiidi kuupäev.

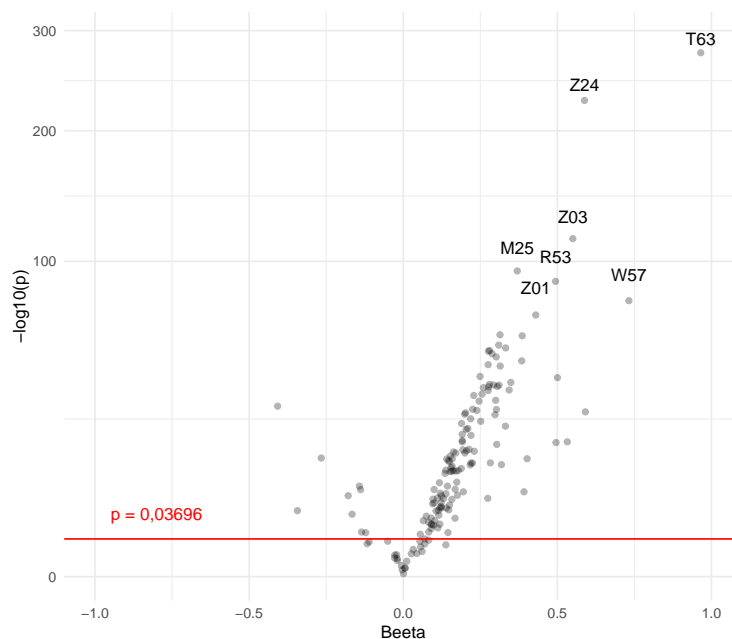
Puukborreliosisiga kaasnevate diagnooside leidmiseks tuli saada kõik puukborreliosisi diagnoosimise kuupäevad ning uurida kõiki diagnoose, mis järgmise 90 päeva jooksul arstide poolt märgiti. Seejärel saab leida, missugused RHK-10 koodid kaasnevad kõige tihedamini puukborreliosisiga.

Kokku leiti 161 haigust, mis kaasnesid puukborreliosisiga 90-päeva jooksul. Andmestikust on eelnevalt eemaldatud kõik RHK-10 koodid, mida esines vähem kui

50 korda. Põhjuseks on see, et neid diagnoose on plaan ka logistilises regressioonis kasutada ning seal on vajalik piisav haigusjuhtumite arv, et hinnata parameetreid. Leidub mitmeid haigusi, kus P53L riskialleeliga inimeste osakaal on tunduvalt suurem võrreldes tavapopulatsiooni jaotusega (Lisa 1). Seal leidub nii kergemaid diagnoose kui ka raskemaid. Üks raske diagnoos, kus on näha P53L alleelidega inimeste suuremat osakaalu, on näiteks krooniline südame isheemiatõbi. Umbes 24% diagnoositutest on kaks riskialleeli pärinud oma vanematelt.

Antud 161 haigusele hinnati ka logistilise regressiooni mudel, hindamaks puukborrelioosi mõju vastava haiguse riskile. Erinevalt eelnevast PheWAS-tüüpi analüüsist on nüüd võetud vanuse mõju täpsemalt arvesse, arvestades ka võimalikku ruutseost.

Logistilise regressiooni mudelis osutus puukborrelioos oluliseks tunnuseks 138 haiguse puhul 161-st (Joonis 8). See viitab sellele, et osa neist seostest võib olla tingitud segavate faktorite poolt.



Joonis 8: Beetakordaja ja negatiivse logaritmitud p-väärtuse hajuvusdiagramm (y-telg ruutjuurega transformeeritud)

## 4.4 Mendeli Randomiseerimine geenivaramu andmetel

Eelnevalt on näha, et puukborrelioosi ja teiste haiguste vahelise seose hindamisel tuleb ette probleem: tuvastatakse selgelt liiga palju olulisi seoseid, mistõttu ei saa selliste analüüside põhjal mõistlikke järeldusi teha. Segavad faktorid võivad siin mängida olulist rolli ja seetõttu on otsustatud kasutada ka Mendeli Randomiseerimist, et hinnata põhjuslikku puukborrelioosi mõju teistele diagnoosidele.

Rakendustarkvara R pakett *MendelianRandomization* võimaldab kasutada Mendeli randomiseerimist andmetel, kus on olemas geneetilisele riskitegurile vastava regressioonikordaja hinnang koos standardvea hinnanguga logistilise regressiooni mudelist mittegeneetilisele riskitegurile ning samuti hinnang sama geneetilise riskiteguri kordajale mudelist haiguse esinemisele.

Meetodit rakendati haigustele, mis kaasnesid puukborrelioosiga 90-päeva jooksul. Eelnevalt oli hinnatud puukborrelioosi regressioonikordaja koos standardveaga. Lisaks hinnati ka P53L geenivariandi regressioonikordajad koos standardveaga, kus kovariaatide hulka lisati ka vanuse ruut.

Uuringus on kasutusel ainult üks geenivariant ning üks mittegeneetiline tegur, seega kõigi 161 haiguse uurimisel jääb nendevaheline beetakordaja ning standardviga samaks (0,1826 ja 0,0129). Kui P53L ja uuritavatate haiguste parameetrid on olemas, siis antakse need andmed MR funktsioonile sisendiks ning rakendatakse meetodit (Tabel 5).

Tabel 5: Mendeli randomiseerimise olulised tulemused Benjamin-Hochbergi korrigeerimisega

RHK-10	Nimetus	$\beta$	SE	P-väärtus
A69	Muud spiroheetnakkused	0,98	0,06	$2,07 \cdot 10^{-55}$
T63	Mürgiste loomade toks. toime	0,57	0,08	$3,46 \cdot 10^{-14}$
W57	Mittemürgiste putukate toime	0,61	0,10	$4,09 \cdot 10^{-09}$
X25	Kokkupuude mürgiste lüliljalgsetega	0,71	0,16	$7,90 \cdot 10^{-06}$
M50	Kaela lülivahekeste haigusseis.	-0,24	0,06	$1,32 \cdot 10^{-04}$

Kui esimesed 4 diagnoosi on positiivsete beetakordajatega, viidates sellele, et puukborrelioos suurendab nende diagnooside saamist, siis viimasega on näha olulist erinevust. Selle põhjal saaks väita, et puukborrelioosi põdemine vähendab kaela lülivahekestate probleeme.

Antud ajahetkel ei ole leitud täpset selgitust MR-ga leitud olulistele seostele. Kuigi on arusaadav, et enamik olulistest seostest põhinevad putukatega kokkupuutel, jääb segaseks kaela lülivahekestate haigusseisundi diagnoos.

Kontrollimiseks tuleb vaadata, kuidas MR on mõjutanud eelnevaid tulemusi. MRi tulemuste mõju haigustele saab hinnata kõige paremini puukborrelioosi beetakordajaid enne ja pärast MRi võrreldes (Tabel 6).

Tabel 6: Mendeli randomiseerimise beetakordajate ja standardvea võrdlus

RHK-10	Diagnoos	Enne MR $\beta$ ja SE	Peale MR $\beta$ ja SE
A69	Muud spiroheetnakkused	23,26 ja 91,22	0,98 ja 0,06
T63	Mürgiste loomade toksiline toime	0,97 ja 0,03	0,57 ja 0,08
W57	Mittemürgiste putukate toime	0,73 ja 0,04	0,61 ja 0,10
X25	Kokkupuude mürgiste lüljalgsetega	1,12 ja 0,05	0,71 ja 0,16
M50	Kaela lülivahekestate haigusseisundid	0,19 ja 0,03	-0,24 ja 0,06

Mendeli randomiseerimine on enamasti vähendanud beetakordajat ehk muutnud puukborrelioosi mõju teistele diagnoosidele väiksemaks ning suurendanud standardviga, mille tõttu on olulisi seoseid alles jäänud 5. Kordaja märgimuutus on toimunud kaela lülivahekestate haigusseisundi puhul, kus MR-ga on beetakordaja negatiivseks muutunud.

Märgierinevus viitab sellele, et eeldused MRi jaoks võivad olla rikutud. P53L ei pruugi olla täiesti sõltumatu uuritavatest haigustest, kui arvestada segavaid fakto-

reid. Võimalus on, et P53L geenivariandil eksisteerib seos segavate faktoritega, mis mõjutavad puukborrelioosi riski või teiste uuritavate haiguste riski.

## Kokkuvõte

Lõputöö eesmärk oli uurida põhjuslikke seoseid puukborrelioosi või puukborrelioosi nakatumist soodustava inimeeni ja teiste haiguste diagnooside vahel. Sooviti leida tulemusi, mida perearstid saaksid kaasuvate haiguste või raskekujulise puukborrelioosi ennetamisel kasutada.

Töös kasutati seoste uurimisel hii-ruut testi ning logistilist regressiooni. Lisaks oli rakendatud seoseuuringu meetodit (PheWAS), mis põhineb samuti logit mudelil. Põhjuslikkuse uurimiseks on kasutusse võetud ka Mendeli randomiseerimine.

Eelnevalt mainitud metoodikat rakendati Tartu Ülikooli Eesti geenivaramu andmetel. Selle jaoks kinnitati esmalt P53L ja puukborrelioosi vaheline seos. Seejärel uuriti P53L ja puukborrelioosi mõju teiste haiguste nakatumisriskile kasutades logistilist regressiooni (ka PheWAS). Lisaks uuriti 90-päeva jooksul peale puukborrelioosi diagnoosi saamist kaasnevaid haigusi ning rakendati logistilist regressiooni nendel andmetel. Lõpuks kasutati tulemuste peal Mendeli randomiseerimist, et segavate faktorite mõju vähendada ning puukborrelioosi põhjuslikku mõju uurida.

Meetodite rakendamisel ei leitud märkimisväärseid tulemusi, mida saaks tervishoius kasutada. Tuli välja, et puukborrelioosi põdenud inimestel on seoseid paljude teiste haigustega. Seal võib olla probleem segavates faktorites, mis hõlmavad näiteks üldiselt inimese eluviisi.

Mendeli randomiseerimise kasutamisel oli näha, et paljud seosed muutusid ebaolulisteks. Lisaks vähenes sellega puukborrelioosi beetakordaja ning suurenes standardviga. Kõik haigused peale ühe olid seotud putukate või lüljalgsetega, mis on arusaadav puukborrelioosi puhul. Üllatuslik oli see, et puukborrelioosi põdenutel on väiksem võimalus saada kaelavigastusi (MR-ga muutus beetakordaja positiivsest negatiivseks). See võib viidata sellele, et puukborrelioosi põdenud inimesed teevad vähem kontoritööd, mis võib soodustada kaelaprobleeme või leitud seos on lihtsalt juhuleid.

Kuigi me ei näe erilisi erinevusi kaasuvate haiguste esinemises, siis rahvatervise vaatepunktist on mõistlik SCGB1D2:P53L geenivariandi esinemist arvesse võtta. Kuna puukide ja borrelia bakterite levik erineb maakonniti, siis võib tekkida olukordi, kus geneetiliselt kõrge riskiga inimesed on keskkonna tõttu väga kõrges nakatumisriskirühmas. Lisaks on antud töö raames näha P53L genotüübi ja putukate hammustuste vahel selget seost. Võimalik, et P53L mõjutab putukate käitumist või naha reaktsiooni putukahammustustele, mida võiks edasi uurida. Lõpuks saab veel uurida FinnGen projekti andmeid, kas seal esineb seos puukborrelioosi ning kaelavigastuste vahel, et töös leitud tulemust valideerida.

SCGB1D2 geeni bioloogiline funktsioon on tundmatu, seega peaks läbi viima uurin-gud, mõistmaks selle geeni toimeainet inimbioloogias. Teades selle geeni olemust, saaks tõhusamini disainida selle mehhanismi mõjutavaid ravimeid ning tulevikus puukborrelioosi levikut pärssida.

## Kasutatud allikad

- Aguero-Rosenfeld, Maria E ja Gary P Wormser (2015). “Lyme disease: diagnostic issues and controversies”. *Expert Review of Molecular Diagnostics* 15.1. PMID: 25482091, lk. 1–4. DOI: [10.1586/14737159.2015.989837](https://doi.org/10.1586/14737159.2015.989837).
- Benjamini, Yoav ja Yosef Hochberg (1995). “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”. *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1, lk. 289–300. DOI: <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- CDC, Centers for Disease Control ja Prevention (10. jaanuar 2022). *Post-Treatment Lyme Disease Syndrome*. URL: <https://www.cdc.gov/lyme/postlds/index.html> (vaadatud 09.05.2023).
- Davey Smith, George ja Gibran Hemani (juuli 2014). “Mendelian randomization: genetic anchors for causal inference in epidemiological studies”. *Human Molecular Genetics* 23.R1, R89–R98. ISSN: 0964-6906. DOI: [10.1093/hmg/ddu328](https://doi.org/10.1093/hmg/ddu328).
- Denny, Joshua C., Marylyn D. Ritchie, Melissa A. Basford, Jill M. Pulley, Lisa Bastarache, Kristin Brown-Gentry, Deede Wang, Dan R. Masys, Dan M. Roden ja Dana C. Crawford (märts 2010). “PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations”. *Bioinformatics* 26.9, lk. 1205–1210. ISSN: 1367-4803. DOI: [10.1093/bioinformatics/btq126](https://doi.org/10.1093/bioinformatics/btq126).
- Huang, Qingyang (2015). “Genetic study of complex diseases in the post-GWAS era”. *Journal of Genetics and Genomics* 42.3, lk. 87–98. DOI: <https://doi.org/10.1016/j.jgg.2015.02.001>.
- Kobayashi, Takaaki, Yvonne Higgins, Michael T. Melia ja Paul G. Auwaerter (2022). “Mistaken Identity: Many Diagnoses are Frequently Misattributed to Lyme Disease”. *The American Journal of Medicine* 135.4, 503–511.e5.

ISSN: 0002-9343. DOI: <https://doi.org/10.1016/j.amjmed.2021.10.040>.

Ku, Chee Seng, En Yun Loy, Yudi Pawitan ja Kee Seng Chia (2010). “The pursuit of genome-wide association studies: where are we now?” *Journal of human genetics* 55.4, lk. 195–206. DOI: <https://doi.org/10.1038/jhg.2010.19>.

Larsson, Susanna C (2021). “Mendelian randomization as a tool for causal inference in human nutrition and metabolism”. *Current opinion in lipidology* 32.1, lk. 1–8. DOI: [10.1097/MOL.0000000000000721](https://doi.org/10.1097/MOL.0000000000000721).

Menyhart, Otília, Boglárka Weltz ja Balázs Gyórfy (juuni 2021). “MultipleTesting.com: A tool for life science researchers for multiple hypothesis testing correction”. *PLOS ONE* 16.6, lk. 1–12. DOI: [10.1371/journal.pone.0245824](https://doi.org/10.1371/journal.pone.0245824).

NIH, National Human Genome Research Institute (4. mai 2023). *Genome-Wide Association Studies*. URL: <https://www.genome.gov/genetics-glossary/Genome-Wide-Association-Studies> (vaadatud 09.05.2023).

Prükk, Tiina, Matti Maimets ja Irja Lutsar (2012). “Eesti Arst”. 91.2. Saabunud toimetusse: 05.03.2012, Avaldamiseks vastu võetud: 06.03.2012, Avaldatud internetis: 31.03.2012, lk. 141–147. URL: <https://eestiarst.ee/lymei-tove-nuudisaegne-diagnostika-ja-ravi/>.

Schlauch, Karen A., Robert W. Read, Gai Elhanan, William J Metcalf, Anthony D. Slonim, Ramsey Aweti, Robert Borkowski ja Joseph J. Grzymiski (2018). “GWAS and PheWAS of Red Blood Cell Components in a Northern Nevadan Cohort”. *bioRxiv*. DOI: [10.1101/475467](https://doi.org/10.1101/475467).

Ścieszka, Joanna, Józefa Dąbek ja Paweł Cieślak (2015). “Post-Lyme disease syndrome”. *Reumatologia/Rheumatology* 53.1, lk. 46–48. DOI: <https://doi.org/10.5114/reum.2015.50557>.

- Sotsiaalministeerium (kuupäev puudub). *RHK-10 andmebaas*. URL: <https://rhk.sm.ee/> (vaadatud 09.05.2023).
- Strausz, Satu, Grace Blacker, Sarah Galloway, Paige Hansen, Samuel E. Jones, Erin Sanders, Nasa Sinnott-Armstrong, FinnGen, Irving L. Weissman, Mark Daly, Tuomas Aivelo, Michal Caspi Tal ja Hanna M. Ollila (2022). “Secretoglobin family 1D member 2 (SCGB1D2) protein inhibits growth of *Borrelia burgdorferi* and affects susceptibility to Lyme disease”. *bioRxiv*. DOI: [10.1101/2022.05.27.493784](https://doi.org/10.1101/2022.05.27.493784).
- Terviseamet (kuupäev puudub). *Puukborrelioos*. URL: <https://www.terviseamet.ee/et/nakkushaigused-a-u/puukborrelioos-ehk-lymei-tobi> (vaadatud 09.05.2023).
- Uffelmann, Emil, Qin Qin Huang, Nchangwi Syntia Munung, Jantina de Vries, Yukinori Okada, Alicia R. Martin, Hilary C. Martin, Tuuli Lappalainen ja Danielle Posthuma (august 2021). “Genome-wide association studies”. *Nature Reviews Methods Primers* 1.1, lk. 59. ISSN: 2662-8449. DOI: [10.1038/s43586-021-00056-9](https://doi.org/10.1038/s43586-021-00056-9).
- Ursinus, Jeanine, Hedwig D Vrijmoeth, Margriet G Harms, Anna D Tulen, Hans Knoop, Stefanie A Gauw, Tizza P Zomer, Albert Wong, Ingrid HM Friesema, Yolande M Vermeeren *et al.* (2021). “Prevalence of persistent symptoms after treatment for lyme borreliosis: A prospective observational cohort study”. *The Lancet Regional Health-Europe* 6, lk. 100–142. DOI: <https://doi.org/10.1016/j.lanepe.2021.100142>.
- VanderWeele, Tyler J., Eric J. Tchetgen Tchetgen, Marilyn Cornelis ja Peter Kraft (mai 2014). “Methodological Challenges in Mendelian Randomization”. *Epidemiology* 25.3, lk. 427–435. DOI: [10.1097/EDE.000000000000081](https://doi.org/10.1097/EDE.000000000000081).

# Lisa 1. Puukborreliosisiga kaasnevate haiguste tabel

ICD10/P53L	Nimetus	P53L 0	P53L 1	P53L 2	Juhud P53L 0	Juhud P53L 1	Juhud P53L 2
N92	Rohke, sage ja korratu menstruatsioon	38,10%	34,92%	26,98%	24	22	17
R03	Ebanormaalsed vererõhünäidud diagnoosita juhul	25,00%	50,00%	25,00%	16	32	16
R21	Lööve ja muud mittespetsiifilised lööbelised nahamuutuse	27,69%	47,69%	24,62%	18	31	16
I25	Krooniline südame isheemiatõbi	28,17%	47,89%	23,94%	40	68	34
R52	Mujal klassifitseerimata valu	25,37%	50,75%	23,88%	17	34	16
L02	Nahaabstsess e -mädanik, -furunkul e -paise ja -karbunkul	32,93%	43,90%	23,17%	27	36	19
Z95	Täpsustamata kanna- ja jalavigastus	27,94%	49,26%	22,79%	38	67	31
M16	Koksartroos e puusaliigese artroos	36,10%	41,46%	22,44%	74	85	46
I20	Stenokardia e rinnaangiin e rinnaahistus	26,23%	51,64%	22,13%	32	63	27
I83	Alajäsemete vaariksud e veenikomud	34,62%	43,27%	22,12%	36	45	23
F06	Muud psüühikahäired ajukahjustusest, -düsfunksioonist v	29,47%	48,42%	22,11%	28	46	21
M47	Spondüloos e lüliläikus	34,88%	43,41%	21,71%	45	56	28
K02	Hambasööbija	32,14%	46,43%	21,43%	36	52	24
A69	Muud spiroheetnakkused	41,43%	37,14%	21,43%	29	26	15
I84	Hemorroidid e päraku veenikomud	37,96%	40,88%	21,17%	52	56	29
Z13	Täpsustamata kanna- ja jalavigastus	35,23%	44,32%	20,45%	62	78	36
M48	Muud spondülopaatiad	39,19%	40,54%	20,27%	29	30	15
M50	Tservikaalsete ja kaela lülivahekettaste haigusseisundid	27,66%	52,13%	20,21%	26	49	19
M13	Muud artriidid	39,52%	40,32%	20,16%	49	50	25
E05	Tür[e]otoksikoos [hüpertüreosis]	35,80%	44,44%	19,75%	29	36	16
M77	Muud entesopaatiad	34,01%	46,26%	19,73%	50	68	29
J04	Äge kõripõletik e larüngiit ja hingetorupõletik e trahheiid	25,00%	55,68%	19,32%	22	49	17
M06	Muud reumatoidartriidid	39,16%	41,57%	19,28%	65	69	32
J01	Äge sinu[s]iit e ninakõrvalurkepõletik	28,33%	52,50%	19,17%	34	63	23
L03	Tselluliit	42,86%	38,10%	19,05%	54	48	24
K04	Hambasäsi ja juuretipuümbrise haigused	47,62%	33,33%	19,05%	30	21	12
C50	Rinna pahaloomuline kasvaja	38,00%	43,00%	19,00%	38	43	19
M80	Patoloogilise murruga osteoporoos e luu-urnemus	43,75%	37,50%	18,75%	28	24	12
L40	Psoriaas e soomussammaspool	44,12%	37,25%	18,63%	45	38	19
H81	Vestibulaarfunktsiooni e esikatalituse häired	32,00%	49,50%	18,50%	64	99	37
L23	Allergiline kontaktdermatiit	33,97%	47,71%	18,32%	89	125	48
R42	Vertiigo e peapööritus	30,37%	51,85%	17,78%	41	70	24

Joonis 9: Esimesed suurima P53L osakaaluga puukborreliosisiga 90 päeva jooksul kaasnevad haigused

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

Mina, Sten Raak,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose "Puukborrelioos: geneetilised riskitegurid ja nakkusega seotud terviseriskid TÜ Eesti geenivaramu andmete põhjal", mille juhendajad on Erik Abner ja Krista Fischer, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 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.

Sten Raak

09.05.2023