

TARTU ÜLIKOOL  
MATEMAATIKA-INFORMAATIKATEADUSKOND  
MATEMAATILISE STATISTIKA INSTITUUT

Marja-Liisa Roos

**Hüpertensiooni geneetilise riskiskoori prognoosivõime hindamine:  
analüüs Tartu Ülikooli Eesti Geenivaramu andmebaasi põhjal**

Magistritöö

Juhendaja Krista Fischer, PhD  
vanemteadur  
TÜ Eesti Geenivaramu

TARTU  
2013

# SISUKORD

SISSEJUHATUS .....	3
1 STATISTILINE METOODIKA .....	5
1.1 Lineaarne regressioon .....	5
1.2 Logistilise regressiooni mudel .....	5
1.3 Diagnostiliste testide analüüsi meetodid .....	6
1.3.1 Tundlikkus ja spetsiifilisus .....	6
1.3.2 Positiivne ja negatiivne prognoosiväärtus .....	6
1.3.3 ROC-kõver .....	7
1.3.4 AUC ja C-statistik .....	10
1.3.5 Ümberklassifitseerimise indeks (NRI) .....	10
2 TÕ EESTI GEENIVARAMU ANDMETE ANALÜÜS .....	16
2.1 Andmestiku kirjeldus .....	16
2.2 ROC-kõverad .....	24
2.3 Lineaarsete mudelite koostamine vererõhu tasemele .....	35
2.4 Alamandmestiku koostamine ja tulemused .....	37
2.5 Ümberklassifitseerimise indeksid .....	39
2.6 Kokkuvõtte erinevate mudelite prognoosivõimest .....	50
2.7 Hüpertensiooni riski hindamine .....	52
KOKKUVÕTE .....	55
SUMMARY .....	57
KASUTATUD KIRJANDUS .....	59

## SISSEJUHATUS

Kõrgenenud vererõhk on probleemiks paljudele inimestele ning samuti on ta üheks oluliseks riskiteguriks südamehaiguste tekkimisel. Käesoleva magistritöö eesmärgiks on uurida ülegenoomsete assotsiatsiooniuuringute põhjal leitud geenimarkerite prognoosivõimet kõrgvererõhutõve riski hindamisel.

Ajakirjas „*Nature*“ avaldati oktoobris 2011 artikkel (Ehret jt. 2011, 103), kus paljusid riike haarav uuring tegi kindlaks 29 geenimarkerit ehk SNP-d (*Single Nucleotide Polymorphism*), mis osutusid seotuks vererõhu tasemega, hüpertensiooni esinemisega, vasaku vatsakese seina paksusega, insuldi ja südame isheemiatõve esinemisega. Kõnealuste geenimarkerite leidmine oli küllaltki komplitseeritud, sest vererõhku mõjutavad paljud tegurid näiteks suitsetamine, kehakaal, füüsiline aktiivsus jne. 200 000 vaatluse all olnud Euroopa inimese hulgas sisaldus ka Eesti geenidonorite andmeid sisaldav valim HYPTEST.

Inimese vererõhk on mõjutatud mitmetest bioloogilistest markeritest ning on tundlik keskkonna mõjutustele. Vererõhk on vajalik selleks, et kanda verd südamest üle kogu keha laiali. Vererõhuks on rõhk, mida veresoontes voolav veri avaldab veresoonte seintele. Hüpertensiooni korral on häiritud arteriaalse rõhu kontrollmehhanism. Süstoli ehk südame kontraktsiooni ajal tekkivat rõhku nimetatakse süstoolseks rõhuks ja diastoli ehk südame lõõgastumise ajal tekkivat rõhku diastoolseks rõhuks. (Tervise Arengu Instituut 2013)

Üle miljardi inimese üle kogu maailma kannatab hüpertensiooni all, see tähendab, et süstoolne vererõhk on üle 140 mmHg või diastoolne vererõhk on üle 90 mmHg. Isegi vähene vererõhu kasv on tugevalt seotud südamehaiguste tekke riski kasvuga. (Ehret jt. 2011, 103)

Leitud markerid annavad uusi teadmisi vererõhu taseme geneetikast ning võivad kaasa aidata südamehaiguste ennetamisel.

Antud magistritöö eesmärgiks on uurida, milliste meetodite abil saab hinnata geenimarkerite võimet eristada kõrge ja madala riskiga indiviide, olles eelnevalt arvesse võtnud muud riskitegurid. Töö metoodika osas on tutvustatud järgmisi statistilise analüüsi meetodeid: lineaarne ja logistiline regressioon, diagnostiliste testide analüüsi meetodeid, kuhu alla kuuluvad testi tundlikkuse ja spetsiifilisuse arvutamine, ROC-analüüs ja ROC-kõvera aluse pindala leidmine ja interpreteerimine ning ümberklassifitseerimise indeksi tutvustamine. Töö praktilises osas on esmalt iseloomustatud Eesti Geenivaramu

geenidoonoreid erinevate tunnuste alusel ning analüüsitud hüpertensiooni diagnoosi esinemise seoseid geneetiliste markerite ning erinevate keskkonnast tulenevate tausttunnustega, nagu näiteks suitsetamine ning toitumisharjumused. Seejärel on uuritud saadud mudelite prognoosivõimet, kasutades ROC-kõveraaid ja ümberklassifitseerimise indekseid. Samuti uuriti ka seda, kas prognoosivõime on erinev erinevates geenidoonorite gruppides, sõltuvalt soost, vanusest ja kehamassiindeksist.

Autori panuseks oli statistilise metoodika valik ja kirjeldamine mitmete erinevate allikate põhjal. Lisaks programmide kirjutamine, millega võimaldati geenidoonorite andmete pealt vastavaid mudeleid ja tabeleid koostada ja analüüsida ning graafikuid joonistada.

Töö kirjutamiseks on kasutatud tekstitöötlusprogrammi Microsoft Word ning programmikoodid on koostatud statistikapaketi R abil.

Autor tänab TÜ Eesti Geenivaramut töös kasutatavate geenidoonorite käitumuslike ja nende isikuandmeid puudutavate andmete kasutamise loa ja juhendaja Krista Fischerit täienduste ning paranduste eest.

# 1 STATISTILINE METOODIKA

## 1.1 Lineaarne regressioon

Lineaarse mudeli korral eeldatakse, et sõltuva tunnuse  $Y$  keskvärtus on avaldatav argumenttunnuste  $X$  lineaarse funktsioonina. Tüüpiline lineaarne mudel esitatakse kujul:

$$Y = X\beta + \varepsilon, \quad (1)$$

kus  $Y$  on  $(n \times 1)$  vektor, mis sisaldab  $n$  vaadeldavat väärtust,  $X$  on  $(n \times p)$  maatriks, mille elemente me teame, sageli nimetatakse maatriksit  $X$  mudeli plaanimaatriksiks;  $\beta$  on tundmatute parameetrite  $(p \times 1)$  vektor ning  $\varepsilon$  on  $(n \times 1)$  vektor, mis sisaldab juhuslikke vigu, kusjuures  $E(\varepsilon|X)=0$ . Klassikalise lineaarse regressioonianalüüsi korral eeldatakse, et juhuslikud vead on normaaljaotusega,  $\varepsilon \sim N(0, \sigma^2)$ . Regressioonanalüüsi eesmärgiks on hinnata parameetrite vektorit  $\beta$  ning hindamiseks kasutatakse enamasti vähimruutude meetodit.

## 1.2 Logistilise regressiooni mudel

Logistilist regressiooni kasutatakse juhul kui funktsioontunnus on binaarne, st uuritav tunnus on Bernoulli jaotusega. Funktsioontunnuse  $Y$  ja argumenttunnuste vektori  $X$  korral on logistilise regressiooni eesmärk hinnata tõenäosusi:

$$\pi(x) = P(Y = 1 | X = x).$$

Logistilise regressiooni mudel on defineeritud kui:

$$\pi(x) = \frac{e^{x\beta}}{1 + e^{x\beta}}. \quad (2)$$

Funktsiooni

$$g(p) = \ln\left(\frac{p}{1-p}\right) \quad (3)$$

nimetatakse logit seosefunktsiooniks. Logit funktsiooni olulisemaks omaduseks on see, et ta teisendab lõigul  $[0,1]$  väärtusi omava tõenäosuse  $\pi(x)$  reaalteljele  $(-\infty, \infty)$ . 'väljalõige Encyclopedia of Biostatistics. Volume 4 (Armitage, Colton 2005, 2870)'

Parameetervektori  $\beta$  hindamiseks kasutatakse suurima tõepära meetodit.

### 1.3 Diagnostiliste testide analüüsi meetodid

#### 1.3.1 Tundlikkus ja spetsiifilisus

Olgu  $H_i$  juhuslik suurus, mille väärtus  $i$ -ndal indiviidil on  $H_i=1$ , kui indiviidil on uuritav haigus ning  $H_i=0$  muidu. Defineerime testi tulemuse  $T$ , mille väärtus  $i$ -ndal indiviidil on  $T_i=1$ , kui testi tulemus on positiivne ning  $T_i=0$ , kui testi tulemus on negatiivne. Spetsiifilisus näitab, kui suur on tõenäosus, et terve isik loetakse terveks,  $i$ -nda indiviidi kohta võib spetsiifilisuse kirja panna järgnevalt (Park, Goo, Jo 2004, 11-12):

$$\text{Spetsiifilisus} = P(T_i = 0 | H_i = 0) \quad (4)$$

ning tundlikkus näitab, kui suure tõenäosusega avastab diagnostiline test haiguse:

$$\text{Tundlikkus} = P(T_i = 1 | H_i = 1). \quad (5)$$

Haiguse levimus, mida tähistame lühendiga PR, tähistab tõenäosust, et juhuslikult populatsioonist valitud inimene on haige ning kasutades eelnevaid tähistusi, võib selle  $i$ -nda indiviidi jaoks kirja panna järgnevalt (Park, Goo, Jo 2004, 11-12):

$$PR = P(H_i = 1). \quad (6)$$

Spetsiifilisus ja tundlikkus on peamised diagnostiliste testide tulemuslikkuse mõõtjad. (Park, Goo, Jo 2004, 11-12)

#### 1.3.2 Positiivne ja negatiivne prognoosiväärtus

Positiivne prognoosiväärtus, mida tähistame lühendiga PPV (*Positive Predictive Value*), näitab kui suur on tõenäosus, et inimene on haige, kui diagnostiline test on positiivne:

$$\begin{aligned} PPV &= P(H_i = 1 | T_i = 1) = \\ &= \frac{PR \cdot P(T_i = 1 | H_i = 1)}{PR \cdot P(T_i = 1 | H_i = 1) + (1 - PR)(1 - P(T_i = 0 | H_i = 0))} \quad (7) \end{aligned}$$

ning negatiivne prognoosiväärtus, mida tähistame lühendiga NPV (*Negative Predictive Value*) näitab, kui suur on tõenäosus, et inimene on terve, kui diagnostiline test on negatiivne (Greenhalgh 1997):

$$\begin{aligned} NPV &= P(H_i = 0 | T_i = 0) = \\ &= \frac{(1 - PR) \cdot P(T_i = 0 | H_i = 0)}{(1 - PR) \cdot P(T_i = 0 | H_i = 0) + (1 - PR)(1 - P(T_i = 1 | H_i = 1))}. \quad (8) \end{aligned}$$

Tõepärasuhe positiivse testi korral (*Likelihood Ratio of a Positive Test*) näitab, kui palju tõenäolisem on positiivne testitulemus haigetel, võrreldes tervetega:

$$LR+ = \frac{P(T_i = 1|H_i = 1)}{1 - P(T_i = 0|H_i = 0)} \quad (9)$$

ning tõenäolisuse negatiivse testi korral näitab, kui palju tõenäolisem on negatiivne testitulemus tervetel, võrreldes haigetega (Greenhalgh 1997):

$$LR- = \frac{1 - P(T_i = 1|H_i = 1)}{P(T_i = 0|H_i = 0)}. \quad (10)$$

### 1.3.3 ROC-kõver

Väga tihti otsustatakse testitulemuse  $T$  üle mõne pideva tunnuse väärtuse põhjal. Kui see tunnus on tähistatud  $M$ -ga, siis  $T_i = I(M_i \geq u)$  ehk testi tulemus loetakse positiivseks, kui tunnuse  $M$  väärtus jääb suuremaks nn otsustuspiirist  $u$ . Mõnikord vastavad positiivsele testi tulemusele tunnuse  $M$  väiksemad väärtused st  $T_i = I(M_i < u)$ . Tunnuse  $M$  kasutatavust testimisel ning testi omadusi erinevate otsustuspiiride korral kirjeldatakse ROC-kõverate (*Receiver Operating Characteristic Curve*) abil.

ROC-kõver on efektiivne meetod hindamaks diagnostiliste testide tulemuslikkust, mis on graafikul defineeritud vastavalt: y-teljel tundlikkus ning x-teljel valepositiivse tulemuse tõenäosus ehk (1-spetsiifilisus). ROC-kõver on erinevatele otsustuspiiridele vastavate tundlikkuse ja spetsiifilisuse paaride graafiline esitus hindamaks prognoosi täpsust. 'väljalõige Encyclopedia of Biostatistics. Volume 7 (Armitage, Colton 2005, 4523)'

Võib öelda, et ROC-kõver hindab, kui hästi test või mudel eraldab üksikjuhud kahte klassi- haiged ning terved. (Cook 2008, 17)

Alljärgnevalt tuuakse näide ROC-kõvera kasutamise kohta:

FEV (*Forced expiratory volume*)- kopsumaht ehk väljahingatava õhu maht pärast sügavat sissehingamist on väga oluline suurus kopsufunktsiooni mõõtmiseks. Olgu FEV1 kopsumaht esimesel sekundil. (Campbell ja Machin 1999, 44). Tulemused on toodud tabelis 1.3.3.1.

**Tabel 1.3.3.1.** FEV1 väärtused (% normaalsest tasemest) inimestel, kellel oli ja kellel polnud pneumokonioosi ehk kopsutolmustust. (Campbell ja Machin 1999, 44)

<b>Mehed, kellel on pneumokonioos (n=27)</b>
40 43 47 49 50 50 53 57 58 58 58 62 65 69 71 73 74 75 75 77 78 79 80 87 90 100 105
<b>Mehed, kellel pole pneumokonioosi (n=13)</b>
60 67 73 75 79 80 83 87 89 100 105 109 115

Andmestikus, mis on toodud tabelis 1.3.3.1 on 40 mittesuitsetaja FEV1 väärtused pneumokonioosi esinemise lõikes. Andmestikus on toodud väärtused protsentides normaalsest tasemest. (Campbell ja Machin 1999, 44)

Diagnostiliste otsuste tegemiseks hingamisteede haiguste korral kasutatakse enamasti otsustuspiirina FEV1 väärtust 80% oodatavast väärtusest terve inimese korral. Tuginedes tabelile 1.3.3.1 saame (2 x 2) sagedustabeli 1.3.3.2, arvestades valitud otsustuspiiri. (Campbell ja Machin 1999, 44)

**Tabel 1.3.3.2.** Sagedustabel pneumokonioosi esinemise lõikes otsustuspiiriga 80%. (Campbell ja Machin 1999, 44)

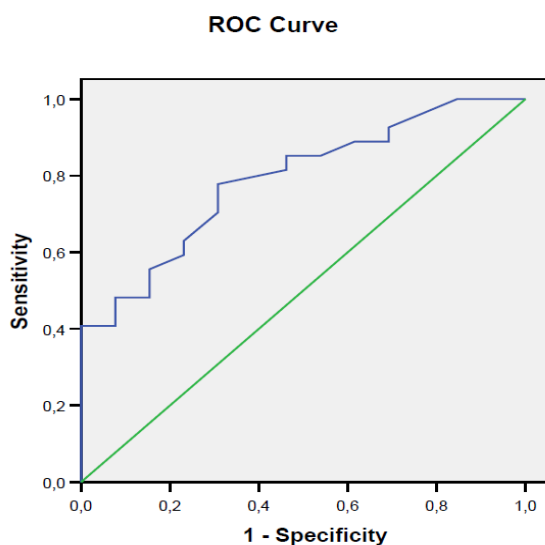
FEV1 väärtus	Haigus ilmnes	Haigus puudus	Kokku
<80% normaalsest tasemest	22	5	27
≥80% normaalsest tasemest	5	8	13
Kokku	27	13	40

Tabelist 1.3.3.2 saame et tundlikkus on  $P(T_i=1|H_i=1)=0.81$  ehk 81% ja spetsiifilisus on  $P(T_i=0|H_i=0)=0.62$  ehk 62%. Siiski ei pea valima FEV1 väärtuseks 80%, teised võimalikud otsustuspiiride valikud selle andmestiku puhul võivad olla vahemikust FEV1=40% kuni FEV1=115%. Igale võimalikule otsustuspiirile on vastavusse seatud spetsiifilisus ja tundlikkus, tulemused on toodud tabelis 1.3.3.3. (Campbell ja Machin 1999, 44)

**Tabel 1.3.3.3.** Tundlikkus ja spetsiifilisus erinevate otsustuspiiride lõikes.

FEV1 väärtus ehk otsustuspiir u	Haigete arv, kellel $M < u$	Tervete arv, kellel $M \geq u$	Spetsiifilisus	Tundlikkus
40%	0	13	0%	100%
50%	4	13	15%	100%
60%	11	13	41%	100%
70%	14	11	52%	85%
80%	22	8	81%	62%
90%	24	4	89%	31%
100%	25	4	93%	31%
110%	27	1	100%	8%
120%	27	0	100%	0%

Joonisel 1.3.3.1 on toodud tabeli 1.3.3.3 põhjal saadud ROC-kõver, kus y-telje väärtused vastavad tundlikkusele ja x-telje väärtused valepositiivse tulemuse tõenäosusele. (Campbell ja Machin 1999, 45)



**Joonis 1.3.3.1.** ROC-kõver kopsutolmustuse andmetele. (Campbell ja Machin 1999, 45)

Täiuslik diagnostiline test on selline, mis ei anna valepositiivseid ja valenegatiivseid tulemusi. Graafiliselt algaks ROC-kõver nullpunktist ning liiguks edasi mööda y-telge, kuni tundlikkus on 1 ning edasi toimuks liikumine paralleelselt x-teljega. Kui test annab valepositiivseid ja õigeid positiivseid tulemusi sama suure tõenäosusega, siis seesugust olukorda iseloomustab sirge  $y=x$ . Kasutuskõlbulike diagnostiliste testi ROC-kõverad jäävad ülemisse vasakpoolsesse kolmnurka. Kui tehakse mitu testi ühe ja sama kliinilise probleemi kohta, siis võrdlusmomendi tekkimiseks on kasulik kanda ROC-kõverad ühele graafikule. (Campbell ja Machin 1999, 45)

Diagnostilised testid on harva 100% täpsed, valepositiivsed ja valenegatiivsed tulemused ilmnevad ikkagi. Test on valiidne, kui valepositiivsete ja valenegatiivsete tulemuste tõenäosused on antud konteksti arvestades piisavalt väikesed. (Park, Goo, Jo 2004, 11-12)

Näiteks sõeluuringutel peab testil olema kõrge tundlikkus, spetsiifilisus võib-olla selle võrra madalam, sest positiivse testitulemuse korral tehakse alati täpsemad uuringud.

### 1.3.4 AUC ja C-statistik

AUC (*The Area Under The Receiver Operating-Characteristic Curve*) on defineeritud kui ROC-kõvera alune pindala. Seesugune definitsioon on ekvivalentne tõenäosusega, et juhuslikult valitud haigest ja ilma haiguseta indiviidist moodustatud paaride korral on haigel tunnuse  $M$  väärtus kõrgem, mis näitab kõrgemat tõenäosust neile, kellel tegelikult uuritav haigus esineb (Pencina jt. 2008, 169):

$$\text{AUC} = P(M_i > M_j | H_i = 1, H_j = 0). \quad (11)$$

Valem (11) saadakse spetsiifilisuse definitsioonist (4): kui  $f$  on juhusliku suuruse  $M$  tihedusfunktsioon iga otsustuspiiri  $u$  korral,  $0 < u < 1$ , siis saab tundlikkuse ning valepositiivse tulemuse tõenäosuse kirja panna järgnevalt (Pencina jt. 2008, 169):

$$\text{Tundlikkus}(u) =: S(u) = P(M > u | H = 1) = \int_u^1 f(m | H = 1) dx \quad (12)$$

$$1 - \text{Spetsiifilisus}(u) =: T(u) = P(M > u | H = 0) = \int_u^1 f(m | H = 0) dx. \quad (13)$$

Paneme tähele, et AUC on graafiku  $S(u)$  vs  $T(u)$  joonealune pindala, mida saab esitada kujul (Pencina jt. 2008, 169):

$$\text{AUC} = \int_0^1 S(u) \frac{d}{du} T(u) du = \int_0^1 S(u) f(u | H = 0) du. \quad (14)$$

Tihti kasutatakse ROC-kõvera aluse pindala jaoks AUC-ga ekvivalentset mõistet C-statistik. (Cook 2008, 18) AUC parandamine, lisades uue markeri, käib lihtsalt- arvutatakse AUC-d, kasutades kahte mudelit, millest üks sisaldab markerit ning teine mitte ning võrreldakse neid, st erinevate testimetodite võrdluseks kasutatakse C-statistikute võrdlemist. (Pencina jt. 2008, 169)

### 1.3.5 Ümberklassifitseerimise indeks (NRI)

Tihti defineeritakse otsustamisel kasutatav tunnus  $M$  lineaarse või logistilise regressioonimudeli põhjal saadud prognoosina. Siin pakub huvi erinevate mudelite põhjal saadud prognooside võrdlemine. Olgu meil 2 võimalikku tunnust, näiteks 2 erineva mudeli põhjal saadud prognoosid,  $M_1$  ja  $M_2$ . Saadud prognoose vaadeldakse eraldi patsientidel, kellel uuritava haiguse diagnoos esines ja kellel mitte ning seejärel koostatakse sagedustabelid ehk ümberklassifitseerimise tabelid, kus prognoosid on jagatud kategooriatesse ning uuritakse indiviide, kes  $M_1$  ja  $M_2$  alusel kuuluvad erinevatesse

kategooriatesse. Sündmus *üles* on defineeritud, kui vastavas grupis on prognoosi  $M_1$  alusel liigitatud madalamasse gruppi kui prognoosi  $M_2$  alusel ning sündmus *alla* on defineeritud kui  $M_1$  alusel liigitatud kõrgemasse gruppi kui prognoosi  $M_2$  alusel. Kui  $M_2$  on parem prognoos kui  $M_1$ , siis ootame, et haigetel on sündmuse *üles* esinemissagedus suurem kui sündmuse *alla* esinemissagedus. Tervetel vastupidi, oodatakse, et sündmuse *alla* esinemissagedus oleks suurem kui sündmuse *üles* esinemissagedus.

Ümberklassifitseerimise indeks ehk NRI (*Net Reclassification Improvement*) keskendub ümberklassifitseerimise tabelitele ning on defineeritud (Pencina, D'Agostino ja Steyeberg 2010, 12):

$$NRI = P(\textit{üles}|H = 1) - P(\textit{alla}|H = 1) + P(\textit{alla}|H = 0) - P(\textit{üles}|H = 0). \quad (15)$$

Kasutades Bayesi valemit, kirjutatakse eelnev võrdus (15) teisel, kuid ekvivalentsel kujul (Pencina, D'Agostino ja Steyeberg 2010, 12):

$$NRI = \frac{P(H = 1|\textit{üles}) \cdot P(\textit{üles}) - P(H = 1|\textit{alla}) \cdot P(\textit{alla})}{P(H = 1)} + \frac{P(H = 0|\textit{alla}) \cdot P(\textit{alla}) - P(H = 0|\textit{üles}) \cdot P(\textit{üles})}{P(H = 0)}. \quad (16)$$

Teades, et  $P(H=0)=1-P(H=1)$  ja  $P(H=0|*)=1-P(H=1|*)$ , kus \* tähistab määratlust *üles* või *alla*, siis saadakse (Pencina, D'Agostino ja Steyeberg 2010, 12):

$$NRI = \frac{P(H = 1|\textit{üles}) \cdot P(\textit{üles}) - P(H = 1|\textit{alla}) \cdot P(\textit{alla})}{P(H = 1)} + \frac{(1 - P(H = 1|\textit{alla})) \cdot P(\textit{alla}) - (1 - P(H = 1|\textit{üles})) \cdot P(\textit{üles})}{1 - P(H = 1)}. \quad (17)$$

Pärast mõningaid lihtsustusi, saab eelnev valem kuju (Pencina, D'Agostino ja Steyeberg 2010, 12):

$$NRI = \frac{(P(H = 1|\textit{üles}) - P(H = 1)) \cdot P(\textit{üles})}{P(H = 1) \cdot (1 - P(H = 1))} + \frac{(P(H = 1) - P(H = 1|\textit{alla})) \cdot P(\textit{alla})}{P(H = 1) \cdot (1 - P(H = 1))}. \quad (18)$$

Seega NRI (15) esitatakse kahe komponendi summana, kus üks komponent sisaldab uuritava haiguse esinemist ja teine mitte esinemist ning mõlemat komponenti kaalutakse proportsionaalselt uuritava haiguse esinemissagedusega. Seesugune tagasiulatuv ehk retrospektiivne interpretatsioon indeksile on soovitatav esmahaigestumise uurimiseks. Ülaltoodud NRI (18) väljendus muudab interpretatsiooni etteulatuvaks ehk prospektiivseks

ning võimaldab kasutada suurt komplekti ümberklassifitseerimise mõisteid. (Pencina, D'Agostino ja Steyeberg 2010, 12)

Selline formuleering NRI-le nõuab ümberklassifitseerimiseks kindlat reeglit ning haiguse esinemismäärade arvutamist neile, kes on ümberklassifitseeritud ülespoole, allapoole ja kogu valimilt. Seda võib interpreteerida kui haiguse esinemismäära mõõtu, mis kasvab, kui ümberklassifitseerimine on toimunud ülespoole ning kahaneb, kui ümberklassifitseerimine on toimunud allapoole. (Pencina, D'Agostino ja Steyeberg 2010, 12)

Olgu  $n$  indiviidide arv,  $n_U$  ümberklassifitseeritud ülespoole liikujad ning  $n_D$  ümberklassifitseeritud allapoole liikujad, siis võrduse (15) saab esitada kujul (Pencina, D'Agostino ja Steyeberg 2010, 13):

$$\text{NRI} = \frac{P(H = 1|üles) \cdot n_U - P(H = 1|alla) \cdot n_D}{n \cdot P(H = 1)} + \frac{(1 - P(H = 1|alla)) \cdot n_D - (1 - P(H = 1|üles)) \cdot n_U}{n \cdot (1 - P(H = 1))}. \quad (19)$$

Lugejas olevad suurused näitavad oodatavaid sündmusi- mitu korda on uuritava haiguse esinemisel on ümberklassifitseerimisel liigutud üles ja alla (esimene lugeja) ning oodatavaid sündmusi, mitu korda on uuritava haiguse mitte esinemisel liigutud ümberklassifitseerimisel allapoole ja ülespoole (teine lugeja). Nimetajas on vastavalt kõikide haiguse esinemise ning mitte esinemise juhud. Võrdus (19) ei sõltu riskikategooriate arvust ega isegi nende olemasolust, kuna eeldatakse, et haiguse esinemistõenäosus ümberklassifitseerimisel ülespoole ja allapoole on ühesugune kõikide indiviidide puhul, kes on samal ümberklassifitseerimisel. Ülaltoodud tulemus on üldine, riskikategooriate olemasolu pole vajalik. (Pencina, D'Agostino ja Steyeberg 2010, 13)

### **1.3.5.1 Kategoriseeritud ümberklassifitseerimise indeks**

NRI definitsioonis pole nõuet, mis ütleks, et riski peaks stratifitseerima kategooriatesse. Ainus tingimus on see, kuidas on defineeritud ülespoole ning allapoole ümberklassifitseerimine. Mõnedes valdkondades peavad riskikategooriad olema kindlalt määratletud, sest patsientide ravi sõltub nendest kategooriatest, näiteks südameveresoonkonnahaiguste ennetus, ülejäänud valdkonnad küll püüavad lahti mõtestada riski kategooriaid. Isegi kui kategooriad on kindlalt paika pandud, siis nende sobitamine erinevatele juhtudele on keeruline, sõltuvalt sellest, kas erinevatel mudelitel on erinevad

esinemismäärad, näiteks raske südame-veresoonkonnahaigus, täielik südame-veresoonkonnahaigus, raske südame isheemia tõbi, täielik südame isheemia tõbi jne. See võib kaasa tuua erinevad NRI väärtused samale markerile, mis on lisatud erinevatesse mudelitesse. (Pencina, D'Agostino ja Steyeberg 2010, 14)

Kokkuvõtlikult võib NRI jagada kaheks versiooniks: üks, kus kategooriad peaks kasutama, kui nad on paika pandud antud valdkonnas ning millel on mõju otsuste tegemisele ning teine, kus kategooriaid pole määratud ning mida saab kasutada universaalselt. (Pencina, D'Agostino ja Steyeberg 2010, 14)

Tutvustame alljärgnevalt tähistusi:

- NRI (0,2) –NRI 2 kategooria jaoks, otsustuspiiriga 0,2
- NRI (0,06;0,2) märgib NRI-d 3 kategooria jaoks, paika pandud katkevuspunktidega 0,06 ning 0,2
- NRI(>0) või ka pidev NRI, mida kasutatakse juhul, kui kategooriad puuduvad

Lisaks *NRI haiguse esinemisel* ja *NRI haiguse puudumisel*, mis esindavad kahte kasulikku komponenti kogu NRI-st, esimene neist arvutab korrektse ümberklassifitseerimise summa üle nende, kellel on uuritav haigus ning teine üle nende, kellel uuritavat haigust ei esinenud. (Pencina, D'Agostino ja Steyeberg 2010, 14)

Soovitatakse tuua välja mõlemad summad NRI täielikuks interpreteerimiseks

$$\begin{aligned} \text{NRI haiguse esinemisel} &= P(\text{üles}|H = 1) - P(\text{alla}|H = 1) = \\ &= \frac{P(H = 1|\text{üles}) \cdot P(\text{üles}) - P(H = 1|\text{alla}) \cdot P(\text{alla})}{P(H = 1)} \quad (20) \end{aligned}$$

$$\begin{aligned} \text{NRI haiguse puudumisel} &= P(\text{alla}|H = 0) - P(\text{üles}|H = 0) = \\ &= \frac{P(H = 0|\text{alla}) \cdot P(\text{alla}) - P(H = 0|\text{üles}) \cdot P(\text{üles})}{P(H = 0)} \quad (21) \end{aligned}$$

ning NRI on haiguse esinemise NRI ja haiguse mitte esinemise NRI summa. Tavaliselt ei soovitata kasutada rohkem, kui kolme kategooriat, välja arvatud, siis kui kategooriad on paika pandud ning on õigustatud vajadus nende kasutamiseks. Kolm kategooriat pakub piisava kategoriseerimise- kõrgem kategooria kõrgema riski jaoks, individid, keda peaks ravima, madalam kategooria madala riski jaoks, individid, kes ei vaja ravi ning keskmine kategooria kõigi teiste jaoks. Kategooriate kasutamine on õigustatud ainult siis, kui ravi kahes kategoorias olevate indiviide jaoks on erinev. (Pencina, D'Agostino ja Steyeberg 2010, 15)

### 1.3.5.2 Pidev ümberklassifitseerimise indeks, seotus teiste tulemustega

Olgu  $Q^{uus}$  haiguse esinemise ennustatud tõenäosus, mis baseerub *uuel* ennustamise algoritmil ning tõenäosus  $Q^{vana}$  põhineb *vanal*, kus *uus* algoritm on saadud pärast ümberklassifitseerimist. Sellisel juhul defineeritakse pidev ümberklassifitseerimise indeks  $NRI(>0)$  suvaliselt valitud haige  $i$ -nda indiviidi ja suvaliselt valitud terve  $j$ -nda indiviidi jaoks järgnevalt (Pencina, D'Agostino ja Steyeberg 2010, 15):

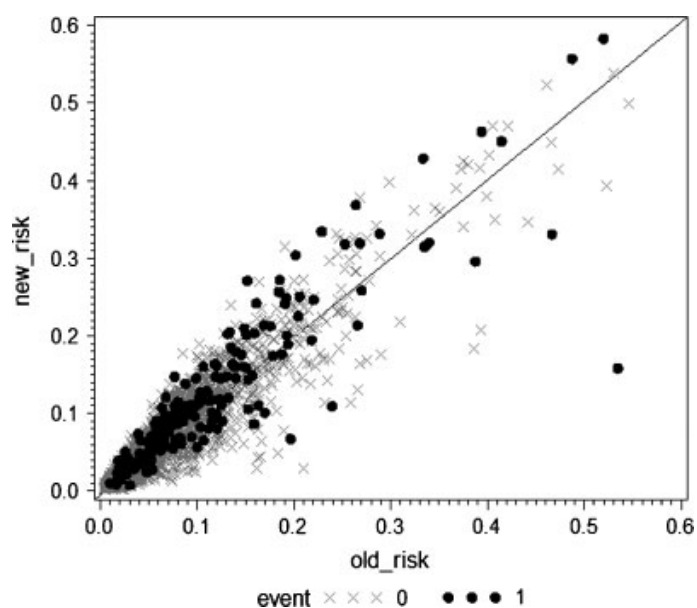
$$NRI(> 0) = P(Q_i^{uus} > Q_i^{vana} | H_i = 1) - P(Q_j^{uus} > Q_j^{vana} | H_j = 0) \quad (22)$$

ning C-statistiku muut avaldub nende jaoks järgmiselt:

$$\Delta C = P(Q_i^{uus} > Q_j^{uus} | H_i = 1, H_j = 0) - P(Q_i^{vana} > Q_j^{vana} | H_i = 1, H_j = 0). \quad (23)$$

$NRI(>0)$  arvutamisel võrreldakse kahte riski ennustamise algoritmi samades sündmuste gruppides, C-statistiku muudus toimub võrdlemine erinevates sündmuste gruppides. Sarnaselt C-statistikule, ei ole ka NRI mõjutatud haiguse esinemise sagedusest, seega on võimalik võrrelda erinevaid uuringuid. (Pencina, D'Agostino ja Steyeberg 2010, 15)

Ümberklassifitseerimist illustreerib joonis 1.3.5.2.1.



**Joonis 1.3.5.2.1.** Ümberklassifitseerimise graafik (Pencina, D'Agostino ja Steyeberg 2010, 15)

Jooniselt 1.3.5.2 on näha, et kategooriate vaba  $NRI(>0)$  saab vaadata kui kokkuvõtlikku suurust ümberklassifitseerimise graafikule. Selle koostamiseks pannakse ühele teljele  $M_2$  ehk *uue* riski ennustamise reegluga arvutatud tõenäosused ja teisele teljele  $M_1$  ehk *vana*

reegliga arvutatud tõenäosused, tähistades haiguse esinemise ja mitte esinemise erinevate sümboolitega. Joon  $y=x$  näitab, et ei ole vahet, kas riski ennustada *vana* või *uue* reegli järgi. Et parandada ümberklassifitseerimise tulemusi oodatakse, et juhud, kus haigus esines oleksid joonest üleval st ennustamise tõenäosus on kasvav ning juhud, kus haigus ei esinenud oleksid joonest allpool st ennustamise tõenäosus on kahanev. (Pencina, D'Agostino ja Steyeberg 2010, 15)

## 2 TÜ EESTI GEENIVARAMU ANDMETE ANALÜÜS

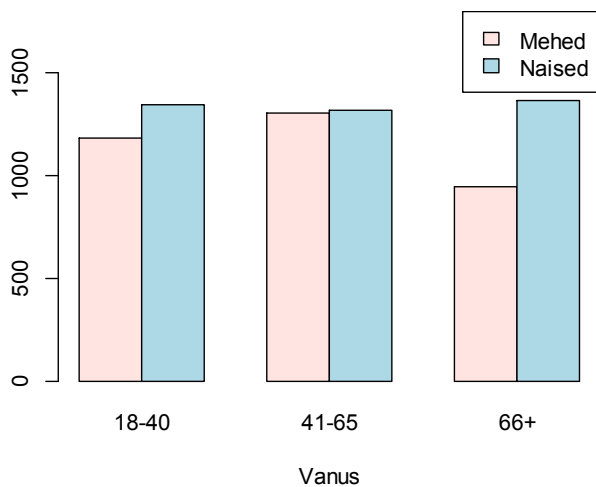
Tartu Ülikooli Eesti Geenivaramu (TÜ EGV) on Eesti rahva tervise-, sugupuu- ja geenandmete kogu, kuhu kuulub ligi 5% Eesti täiskasvanud elanikkonnast. Biopankadesse kogutud andmed võimaldavad läbi viia teadusuuringuid, mis aitavad mõista, kuidas geneetiline informatsioon ja selle muutused mõjutavad üksikisikuid, nende arengut, vananemist, heaolu ja haigusi, aga ka inimeste käitumist ja psühholoogiat seoses muutustega, mis on tingitud neid ümbritsevast keskkonnast ja inimeste elustiilist. TÜ EGV eesmärgiks on rakendada teadustöö tulemusi igapäevaellu, et seeläbi aidata kaasa personaalse meditsiini arengule Eestis. Geenivaramu andmebaas, koosneb 51 535 (seisuga 9. veebruar 2013) geenidoonori andmetest, kellel vanus alates 18-ndast eluaastast. (Eesti Geenivaramu 2013)

### 2.1 Andmestiku kirjeldus

Käesolevas magistritöös on vaatluse all 7471 genotüüpiseeritud geenidoonorit, kellel on määratud üle 700 000 geenimarkeri ehk SNP-i Illumina Omni Express kiibi abil. Neist 38,1% on saanud, kas hüpertooniatõve diagnoosi geenivaramuga liitumise hetkeks või on neil tuvastatud kõrgeenenud vererõhk: süstoolne  $\geq 140$  mmHg või diastoolne  $\geq 90$  mmHg. Antud töös on kasutatud binaarset tunnust, mille väärtus 1 näitab, et doonoril on diagnoositud hüpertensioon ning 0 näitab, et mitte.

Valim on vanuse järgi jaotatud kolmeks vanusegrupiks: 18-40aastased, 41-65aastased ning vanemad kui 66aastased valimisse kuulunud geenidoonorid. Vaadeldakse, kuidas valim jaotub sooliselt ning vanuseliselt.

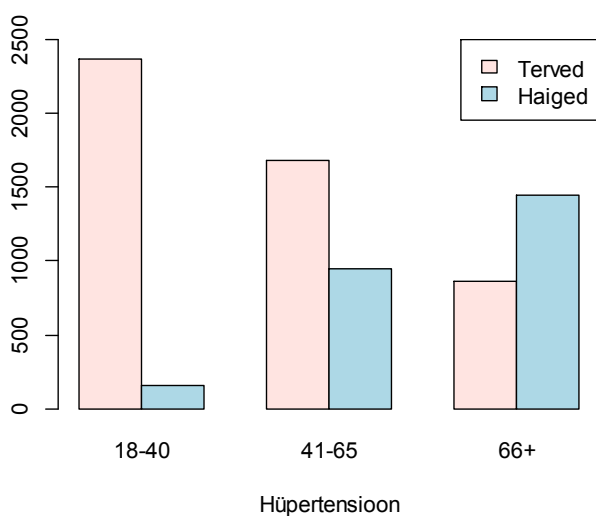
Uurimise all on 3435 meest ning 4036 naist, kes jagunevad kolme vanusegrupi vahel järgnevalt: 18-40aastaste hulka kuulub 1182 meest ja 1350 naist, 41-65aastaste gruppi kuulub 1304 meest ning 1321 naist ning vanusegruppi 66aastased ja vanemad kuulub 949 meest ning 1365 naist. Graafiline illustratsioon on toodud joonisel 2.1.1.



**Joonis 2.1.1.** Vaatluse all olnud geenidoonorite vanuseline ja sooline koosseis.

Uuritavate rahvuseline koosseis on järgmine: 7012 eestlast ning 459 muulast. Vaadeldes geenidoonorite suitsetamisharjumisi on näha, et andmestikus on 2097 praegust suitsetajat, 1197 endist suitsetajat, 4170 mittesuitsetajat ning 7 geenidoonorit, kelle suitsetamisharjumused on teadmata.

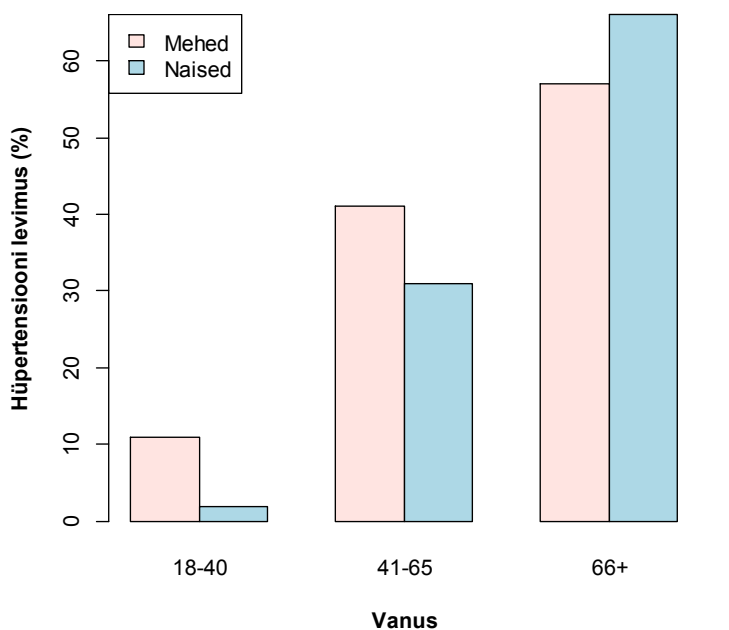
Veel huvituti, kuidas jaguneb hüpertensiooni esinemine vanuselisel. Selle tarvis on toodud joonis 2.1.2.



**Joonis 2.1.2.** Geenidoonorite haigestumine hüpertensiooni erinevates vanusegruppides.

Selgub, et vanuse kasvamisel suureneb hüpertooniatõve diagnoosi esinemine. Vanusegrupis 18-40aastased on hüpertensiooni diagnoosi saanud 160 geenidoonorit, vanusegrupis 41-65aastased on vastava diagnoosi saanud 947 inimest ning 66aastaste ja vanemate seas on kõrgenenud vererõhuga geenidoonoreid 1448. Diagnoosita geenidoonoreid on vastavalt gruppidele 2372, 1678 ning 866 indiviidi.

Joonisel 2.1.3 tuuakse välja hüpertensiooni levimus vanuse ning soo lõikes.

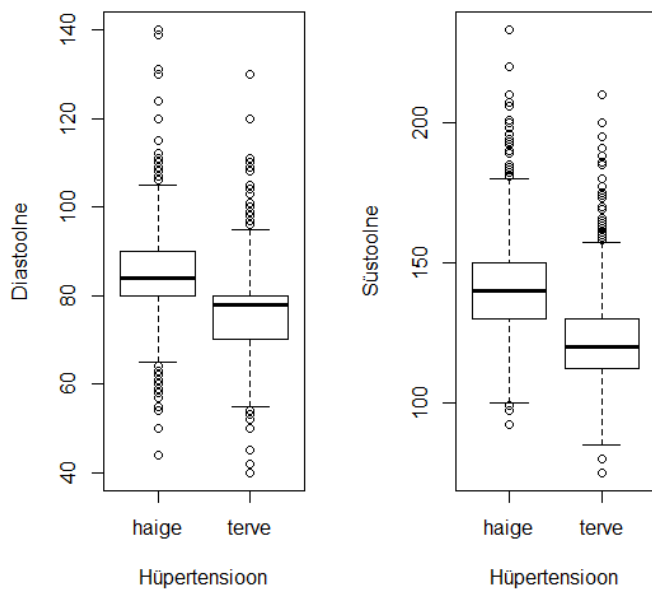


**Joonis 2.1.3.** Hüpertensiooni levimus sõltuvalt soost ja vanusest.

Jooniselt selgub, et nooremas ja keskmises vanusegrupis on hüpertensiooni levimus meestel suurem kui naistel, kuid 66aastaste ja vanemate geenidoonorite hulgas on naistel hüpertensiooni levimus suurem kui meestel. Noorimas vanusegrupis on hüpertensiooni diagnoosiga 128 meest ja 32 naist, keskmises vanusegrupis vastavalt 536 meest ja 411 naist ning 66aastaste ja vanemate hulgas on hüpertensioon diagnoositud 541 mehel ja 907 naisel.

Vaadeldi karpdiagramme süstoolse ja diastoolse vererõhku tasemest hüpertensiooni diagnoosi saanud ja mitte saanud geenidoonoritel. Karpdiagrammide joonistamisel on vaatluse alt välja jäetud 2 geenidoonori andmed, kellede puhul tekkis kahtlus, et tegemist võib olla andmete sisestusveaga. Neist ühe, hüpertensiooni diagnoosi saanud geenidoonori

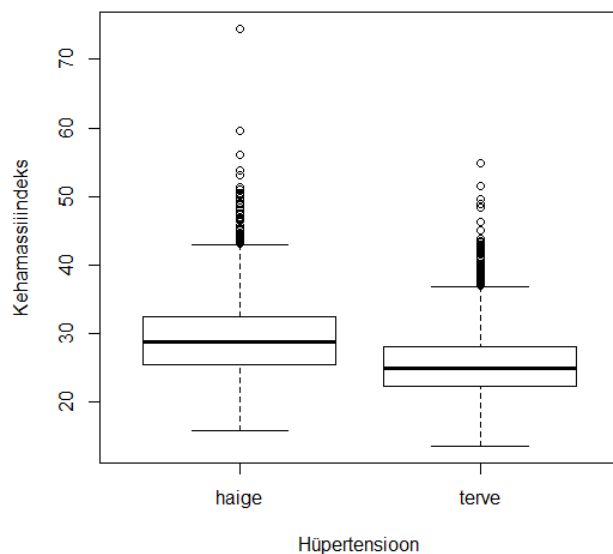
diastoolne vererõhk oli 10 mmHg ning teise hüpertensiooni diagnoosi mitte saanud geenidoonori süstoolne vererõhk oli 17 mmHg.



**Joonis 2.1.4.** Diastoolne ja süstoolne vererõhk hüpertensiooni diagnoosi saanud ja diagnoosi mitte saanud geenidoonoritel, kes kuulusid valimisse.

Ülaltoodud karpdiagrammidel on näha, et kõrgeenenud vererõhu kriteeriumid peavad kaantud valimi puhul paika. Kui tervetel geenidoonoritel on diastoolse vererõhu mediaan 78 mmHg, siis haigetel on see 84 mmHg. Sarnaselt on süstoolse vererõhu mediaan diagnoosi mitte saanud geenidoonoritel madalam, kui hüpertensiooni diagnoosi saanud patsientidel, vastavalt 120 ja 140 mmHg.

Üks hüpertensiooni riskifaktoriks on liigne kehakaal, geenidoonoritel oli arvatud kehamassiindeks, mis avaldub kaalu ja pikkuse ruudu suhtena ning mida mõõdetakse inimese rasvumise määramiseks ja südame-veresoonkonna haigustesse haigestumise riski hindamiseks.

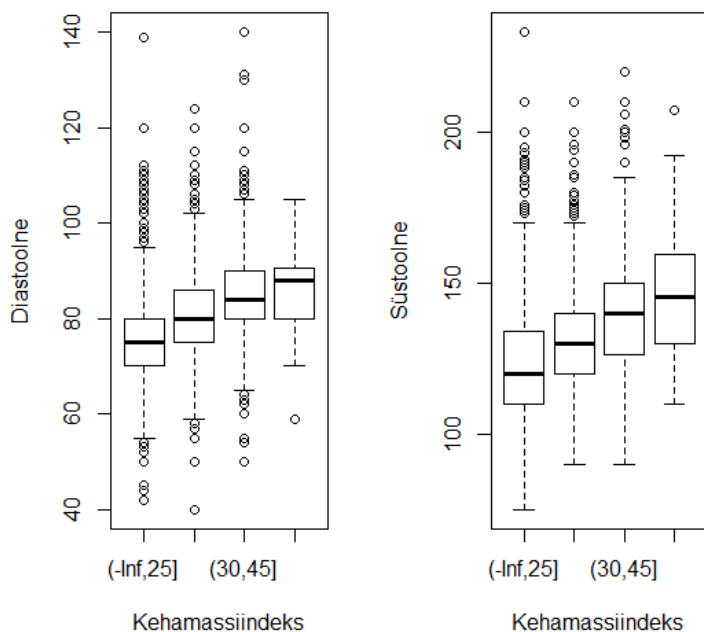


**Joonis 2.1.5.** Hüpertensiooni diagnoosi saanud ja tervete geenidoonorite kehamassiindeks.

Jooniselt 2.1.5 selgub, et hüpertensiooni diagnoosi saanutel on kehamassiindeksi mediaan 28,7 ning diagnoosi mittesaanutel on see 24,8. Kehamassiindeks üle 25 tähistab ülekaalu.

Kõrgenenud vererõhu üheks põhjuseks on liigne kehakaal. Joonistati karpdiagramme, et vaadelda, millisel tasemel on diastoolne ja süstoolne vererõhk erinevate kehamassiindeksite väärtuste vahemike korral. Selle tarvis jagati kehamassiindeksid gruppidesse: esimene grupp, kuhu kuuluvad valimisse kuulunud doonorid kehamassiindeksiga kuni 25 k.a, mis tähistab normaalkaalugruppi, grupp (25;30], mis tähistab ülekaaluliste grupi ning millega kaasneb risk haigestuda diabeeti ja südame-veresoonkonna haigustesse, grupp (30;45], mis on rasvunute grupp, kellel on väga suur risk haigestuda südame-veresoonkonna haigustesse, kõrgvererõhutõvesse ning diabeeti ning viimane grupp üle 45, mida nimetatakse patoloogiliseks ülekaaluks, kellel on võrreldes eelneva grupiga veel kõrgem risk haigestuda eelnimetatud haigustesse.

Tulemused on toodud joonisel 2.1.6. On näha, et kehamassiindeksi suurenemisel kasvab nii diastoolne kui ka süstoolne vererõhk. Vaatluse alt jäid välja eelkirjeldatud 2 patsienti, kellede puhul on ilmselt tegemist sisestusveaga.



**Joonis 2.1.6.** Diastoolne ja süstoolne vererõhk erinevate kehamassiindeksite gruppides.

Toome välja erinevate gruppide korral diastoolse ja süstoolse vererõhu mediaanid: normaalkaalus valimisse kuulunud geenidoonorite diastoolse vererõhu mediaan oli 75 mmHg ning süstoolse 120 mmHg, mida peetakse optimaalseks vererõhuks. Teise kaalugruppi kuuluvate doonorite korral oli diastoolse vererõhu mediaan 80 mmHg ning süstoolne 130 mmHg, mida peetakse normaalseks vererõhuks. Rasvunute grupi korral oli diastoolse vererõhu mediaan 84 mmHg ning süstoolse vererõhu mediaan 140 mmHg, antud juhul on tegemist kerge hüpertensiooniga ning patoloogilises ülekaalugrupis oli vastavad näitajad 88 ja 145,5 mmHg, mille korral on samuti tegemist hüpertensioonitõve diagnoosi saanud geenidoonoritega.

Hüpertooniatõve üheks riskifaktoriks on suitsetamine. Suitsetajatest 428 oli hüpertensiooni diagnoosi saanud, 1669 mitte, endistest suitsetajast kannatas kõrge vererõhu all 568 valimisse kuulunud geenidoonorit, 629 mitte. Mittesuitsetajatest põdes hüpertensiooni 1556 ning ei põdenud 2614 valimisse kuulunud geenidoonorit, kolm doonorit oli diagnoosi saanud, kuid kellele suitsetamisharjumused olid teadmata ning neli geenidoonorit, kes diagnoosi polnud saanud, ei olnud samuti vastanud suitsetamist puudutavatele küsimustele ankeedis. Tabelis 2.1.1 on toodud protsentuaalselt, kuidas hüpertensiooni diagnoos jaguneb erinevate suitsetamisharjumuste korral.

**Tabel 2.1.1.** Hüpertensiooni esinemine geenidoonorite suitsetamisharjumuste lõikes.

		<b>suitsetajad</b>	<b>endised suitsetajad</b>	<b>mittesuitsetajad</b>	<b>suitsetamisharjumused teadmata</b>
<b>Hüpertensiooni diagnoos</b>	Jah	20,41%	47,45%	37,31%	42,86%
	Ei	79,59%	52,55%	62,69%	57,14%

Asjaolu miks endiste suitsetajate grupis jagunevad hüpertensiooni diagnoosi saanud ja mittesaanud geenidoonorid pooleks võib põhjendada sellega, et ilmselt pärast diagnoosi saamist on inimesed loobunud suitsetamisest.

Peale hüpertensiooni uuriti veel geenidoonoritelt, kas nad on saanud südameinfarkti või südame isheemiatõve diagnoosi. Isheemiatõve korral on tegemist südamelihase verevarustuse vähenemise või lakkamisega, mille riskifaktorite hulka kuulub ka hüpertensioon. Südame isheemiatõve raskemaks vormiks on äge südame isheemiatõbi, mida kutsutakse ka südameinfarktiks.

Vaadeldi ka seda, kuidas erineb tõsisemate südamehaiguste nagu südameinfarkt ja südame isheemiatõbi esinemissagedus hüpertensiooni diagnoosiga ja ilma selleta inimestel. Tulemused on toodud tabelis 2.1.2.

**Tabel 2.1.2.** Südameinfarkti ja südame isheemiatõve esinemine sõltuvalt hüpertensiooni diagnoosist.

		<b>Hüpertensiooni diagnoos</b>	
		Jah	Ei
<b>Südameinfarkt</b>	Jah	7,91%	1,85%
	Ei	92,09%	98,15%
<b>Südame isheemiatõbi</b>	Jah	31,74%	8,08%
	Ei	68,26%	91,92%

Südameinfarkt esines 293-l geenidoonoril, kellest hüpertensiooni diagnoosi oli saanud 202. 7178 geenidoonoril südameinfarkti ei esinenud, kuid nendest kannatas hüpertensiooni all 2353 indiviidi. Hüpertensiooni ja südameinfarkti diagnoosi ei saanud 4825 patsienti. Südame isheemiatõbi esines 1208 patsiendil, kellest hüpertensiooni põdes 811. Südame isheemiatõve diagnoosi pole saanud 6263 patsienti, kellest 1744 põeb hüpertensiooni. Geenidoonoreid, kellel ei esine hüpertensiooni ega südame isheemiatõve on valimis 4519.

Autoril oli kasutada andmed ka geenidoonorite hüpertensiooniravimite tarbimise kohta. Ravimi tarvitamise kohta puudust informatsioon kuuel patsiendil, kellest viiel puudus hüpertooniatõve diagnoos. Tulemused on toodud tabelis 2.1.3.

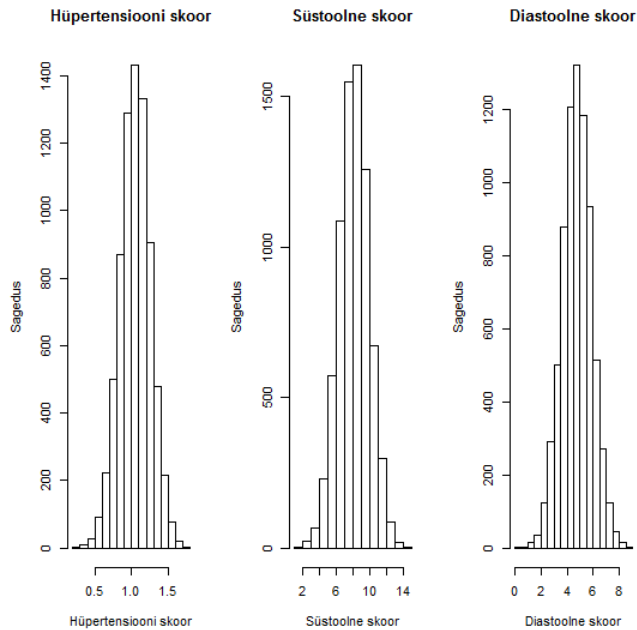
**Tabel 2.1.3.** Hüpertensiooni ravimi tarvitamine diagnoosiga ja diagnoosita geenidoonoritel.

		Hüpertensiooni diagnoos	
		Jah	Ei
Hüpertensiooni ravim	Jah	86,14%	7,11%
	Ei	13,82%	92,79%
	Teadmata	0,04%	0,10%

Geenidoonoritel, kellel diagnoos puudus tarvitas hüpertensiooni ravimit 304 inimest ning ravimit ei tarvitanud 4607 inimest. Geenidoonoritel, kellel esines hüpertooniatõve diagnoos, tarvitas vastavat ravimit 2201 inimest ja ei tarvitanud 353 inimest. Hüpertensiooni ravimi tarvitamine oli teadmata 1 diagnoosiga ja 5 diagnoosita geenidoonori kohta. Diagnoosi olemasolu ei pruugi igal juhul tähendada ravimi tarvitamist, teinekord ei ole haiguse kulg nii kiire, et ravimite manustamine oleks vajalik või tarvitatakse ravimeid, millel võib olla sama toimeaine, mõne muu haiguse leevendamiseks.

Autoril oli kasutada 28 eraldiseisva geenimarkeri andmed valimisse kuulunud geenidoonorite kohta, mille põhjal oli iga indiviidi kohta arvutatud kolm skoori. Iga skoor saadakse, kui korrutada markeri väärtus ehk riskialleelide arv vastava beeta kordajaga artiklist „*Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk*“ (Ehret jt. 2011, 105) ning liidetakse saadud väärtused kokku. Skoorid vastavad diastoolsele ja süstoolsele vererõhule ning hüpertensiooni esinemisele. Joonisel 2.1.7 tuuakse välja kolme riskiskoori histogrammid.

Vaadeldes joonist 2.1.7 on näha, et üle 90% doonorite korral jääb hüpertensiooni skoor vahemikku 0,5-st 1,5ni, seega kasutades hüpertensiooni skoori logistilise regressiooni mudeli parameetrina, võrreldakse erinevusi äärmuslike hüpertensiooni skooriga indiviidide vahel. Kasutades süstoolse vererõhu ja diastoolse vererõhu skoori lineaarse regressiooni mudeli parameetrina, analüüsitakse kui palju erineb keskmine vererõhk skooriühiku kohta.



**Joonis 2.1.7.** Riskiskooride histogrammid.

## 2.2 ROC-kõverad

Kõrgvererõhutõve diagnoosi esinemise uurimisel otsustati keskenduda peamiste teadaolevate riskitegurite uurimisele, milleks on patsiendi sugu, vanus, kehakaal, suitsetamisharjumused ning pärilik eelsoodumus. Selle tarvis koostati logistilise regressiooni mudelid erinevate tingimustega gruppidele, kus esmalt uuriti grupe nii, et pärilikku eelsoodumust ei võetud arvesse. Pärast mudelite koostamist joonistati ROC-kõverad, et eristada päriliku eelsoodumuse mõju. Ilmneb, et suitsetamine ei osutu ühegi grupi korral oluliseks teguriks hüpertensiooni diagnoosi esinemise seisukohast. Ülejäänud hinnangud parameetritele ning šansside suhted kõikidele geenidoonoritele on toodud järgnevas tabelis. Otsustuste tegemiseks on käesolevas magistritöös kasutatud olulisusnivood 0,05.

**Mudel 2.2.1.** Hüpertensiooni esinemine sõltuvalt vanusest, soost ja kehamassiindeksist. Geeniantmeteta mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-8,09	0,22		<0,001
Vanus	0,07	0,002	1,07	<0,001
Naine	-0,35	0,06	0,7	<0,001
BMI	0,14	0,006	1,15	<0,001

Aasta võrra vanematel geenidoonoritel on 1,07 korda suurem šanss saada hüpertensiooni diagnoos (usaldusvahemikuga: 1,064; 1,072). Naistel võrreldes meestega on 30% väiksem šanss saada hüpertooniatõve diagnoos (usaldusvahemikuga: (0,624; 0,790)). Geenidoonoritel, kelle kehamassiindeks on ühe ühiku võrra kõrgem, on 1,15 korda suurem šanss saada hüpertensiooni diagnoos (usaldusvahemikuga: (1,139;1,166)).

Seejärel koostati uus mudel, mis võttis arvesse ka kolme skoori mõju hüpertensiooni esinemisele, et näha, kuivõrd oluline mõju on pärilikul eelsoodumusel hüpertooniatõve diagnoosi esinemisele.

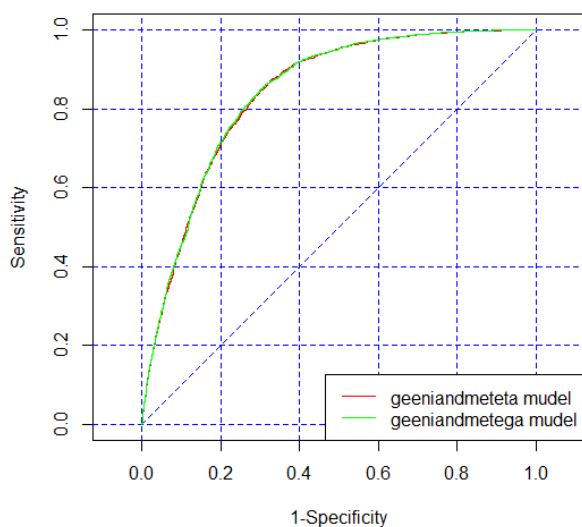
Koostatud mudelis ei osutunud oluliseks diastoolsele ja süstoolsele vererõhule koostatud skooride väärtused, samuti suitsetamine, need tunnused jäeti lõpliku mudeli koostamisel välja. Hinnangud parameetritele ja šansside suhted on toodud järgnevas tabelis.

**Mudel 2.2.2.** Hüpertensiooni esinemine sõltuvalt soost, vanusest, kehamassiindeksist ja ning hüpertensiooni skoori väärtusest. Geeniantmetega mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-8,85	0,276		<0,001
Vanus	0,07	0,002	1,07	<0,001
Naine	-0,36	0,06	0,7	<0,001
BMI	0,14	0,006	1,15	<0,001
Hüpertensiooni skoor	0,68	0,144	1,98	<0,001

Aasta võrra vanematal geenidoonoritel on 1,07 korda suurem šanss saada hüpertooniatõve diagnoos (usaldusvahemikuga: (1,065;1,072)). Naistel võrreldes meestega on 30% (usaldusvahemikuga: (0,622;0,787)) väiksem šanss saada hüpertensiooni diagnoos. Ühe ühiku võrra kõrgemaga kehamassiindeksiga geenidoonorite on 1,15 korda (usaldusvahemikuga: (1,140;1,168)) suurem šanss saada hüpertensiooni diagnoos ning doonoritel, kellel on 1 ühiku võrra kõrgem hüpertensiooni skoor on 1,98 korda suurem (usaldusvahemikuga: 1,494;2,631) šanss saada hüpertooniatõve diagnoos.

Mudelitele 2.2.1 ja 2.2.2 joonistati ROC-kõverad, et hinnata, kui hästi eristub geneetilise eelsoodumuse mõju hüpertensiooni diagnoosi esinemise seisukohalt. Joonisel 2.2.1 on toodud punase joonega mudelile 2.2.1 joonistatud ROC-kõver ning roheline joonega mudelile 2.2.2 joonistatud ROC-kõver. Joonise diagonaal vastab olukorrale, kus argumendi väärtusest sõltumata on tundlikkus ja spetsiifilisus võrdsed 0,5-ga, see tähendab, et hüpertooniatõve diagnoosi saamine on juhuslik ning ei sõltu argumentide väärtustest.



**Joonis 2.2.1.** Mudelitele 2.2.1 ja 2.2.2 joonistatud ROC-kõverad.

Mudelile 2.2.1 joonistatud kõveraalune pindala on 0,84 ning mudelile 2.2.2 joonistatud kõveraalune pindala on 0,842. Nagu jooniselt 2.2.1 on näha langevad nad paljuski kokku. Kuna kõveraalune pindala on mõlemal juhul üle 0,8 tähendab, et mõlema mudeli prognoosivõime on hea. Võib öelda, et mudelid 2.2.1 ja 2.2.2 sobivad küllaltki hästi hindamiseks hüpertooniatõve diagnoosi esinemist, kasutades esimese mudeli korral doonorite sugu, vanust ja kehamassiindeksi väärtust ning teise mudeli korral kasutakse prognoosimisel lisaks arvutatud hüpertensiooni skoori. Tekib küsimus, et kas geeniandmed ei annagi uut infot hüpertensiooni prognoosimiseks, tegelikult peitub vastus selles, et antud juhul on vanuse, soo ja kehamassiindeksi mõju lihtsalt niivõrd tugev, et peidab geneetilise informatsiooni mõju.

Järgnevalt koostati geeniandmetega ja geeniandmeteta mudel eraldi meestele ja naistele, et selgitada välja, kas geeniandmetel on mõju hüpertensiooni diagnoosi esinemisele, hinnangud parameetritele ning šansside suhtele on toodud tabelites 2.2.3 kuni 2.2.6.

**Mudel 2.2.3.** Hüpertensiooni diagnoosi esinemine meestel sõltuvalt vanusest ja kehamassiindeksist. Geeniandmeteta mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-7,90	0,322		<0,001
Vanus	0,05	0,002	1,055	<0,001

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
BMI	0,16	0,010	1,17	<0,001

Mudelitest selgub, et aasta võrra vanematel meesgeenidonoritel on 1,055 korda suurem šanss saada hüpertooniatõve diagnoos (usaldusvahemikuga: (1,052;1,057)) ning 1 ühiku võrra suurema kehamassiindeksiga meesdoonoritel on 1,17 korda suurem šanss saada hüpertensiooni diagnoos (usaldusvahemikuga: (1,164;1,186)).

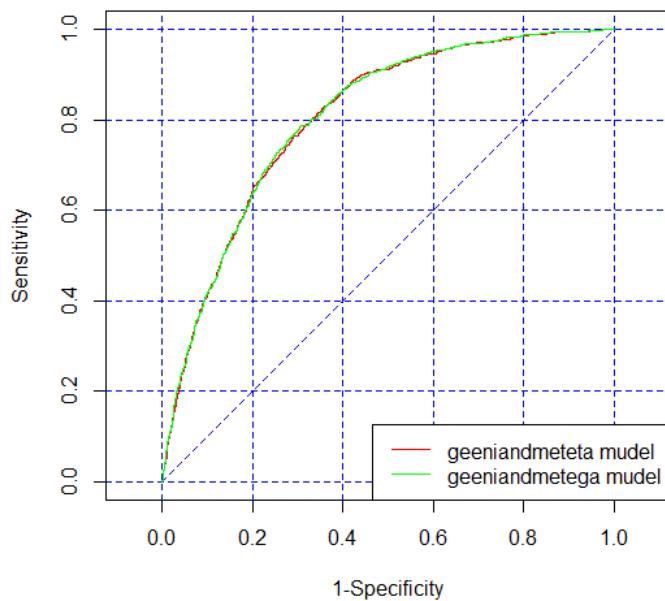
Tuuakse välja ka mudel, kus võetakse arvesse geenianndmed, et näha, kas pärilikul eelsoodumusel on hüpertooniatõve prognoosimisel mõju meessoost geenidonorite seas.

**Mudel 2.2.4.** Hüpertensiooni diagnoosi esinemine meestel sõltuvalt vanusest, kehamassiindeksist ning hüpertensiooni skoori väärtusest. Geenianndmetega mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-8,6	0,4		<0,001
Vanus	0,05	0,002	1,055	<0,001
BMI	0,16	0,01	1,17	<0,001
Hüpertensiooni skoor	0,63	0,2	1,87	0,00166

Tabelist selgub, et meestel vanuse kasvades 1 aasta võrra suureneb hüpertensiooni diagnoosi saamise šanss 1,055 korda (usaldusvahemikuga: (1,053;1,058)), kehamassiindeksi kasvamisest meestel 1 ühiku võrra suureneb šanss saada hüpertooniatõve diagnoos 1,17 korda (usaldusvahemikuga: (1,165;1,187)) ning hüpertensiooni skoori suurenemisel ühe ühiku võrra suureneb šanss saada hüpertensiooni diagnoos 1,87 korda (usaldusvahemikuga: (1,533;2,282)).

Tuuakse välja meeste geenianndmeteta ja geenianndmetega mudelite põhjal joonistatud ROC-kõverad, mis on toodud joonisel 2.2.2.



**Joonis 2.2.2.** Mudelitele 2.2.3 ja 2.2.4 joonistatud ROC-kõverad.

Geeniandmeteta mudelile joonistatud kõveraalune pindala on 0,805 ning geeniandmetega mudelile joonistatud kõveraalune pindala on 0,806. Kõverad langevad kokku ning ka meeste puhul ei ole võimalik eristada geeniandmete mõju hüpertooniatõve diagnoosi prognoosimisel, kuna vanuse ja kehamassiindeksi mõju on niivõrd tugev.

Sarnased mudelid tuuakse ka naiste kohta.

**Mudel 2.2.5.** Hüpertensiooni diagnoosi esinemine naistel sõltuvalt vanusest ja kehamassiindeksist. Geeniandmeteta mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-8,92	0,319		<0,001
Vanus	0,08	0,003	1,08	<0,001
BMI	0,13	0,008	1,14	<0,001

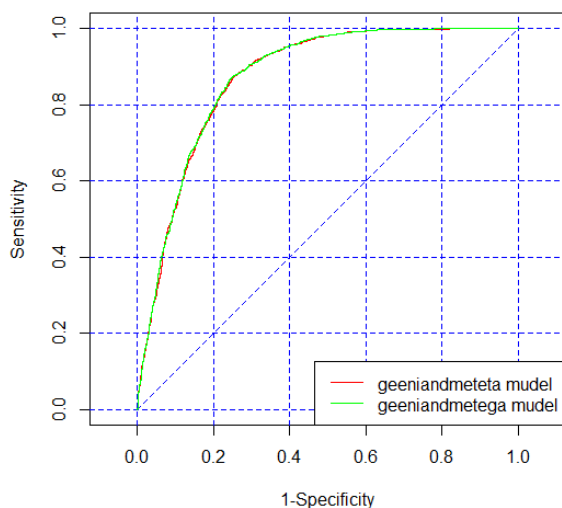
Vaadeldes geeniandmeteta mudelit on näha, et aasta võrra vanematel naisgeenidonoritel on 1,08 korda suurem šanss saada hüpertooniatõve diagnoos (usaldusvahemikuga: (1,079;1,085)) ning ühe ühiku võrra suurema kehamassiindeksiga naisdoonoritel on 1,14 korda suurem šanss saada hüpertensiooni diagnoos (usaldusvahemikuga: (1,128;1,147)).

**Mudel 2.2.6.** Hüpertensiooni diagnoosi esinemine naistel sõltuvalt vanusest, kehamassiindeksist ning hüpertensiooni skoori väärtusest. Geeniandmetega mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-9,72	0,4		<0,001
Vanus	0,08	0,003	1,08	<0,001
BMI	0,13	0,008	1,14	<0,001
Hüpertensiooni skoor	0,73	0,212	2,08	0,00053

Tabelist on näha, et naistel vanuse kasvades 1 aasta võrra on 1,08 (usaldusvahemikuga: (1,079;1,086)) korda suurem šans saada hüpertooniatõve diagnoosi. Samuti suurendab kehamassiindeksi kasvamine 1 ühiku võrra 1,14 (usaldusvahemikuga: (1,129;1,148)) korda šanssi saada hüpertensiooni diagnoos. Naistel, kellel on 1 ühiku võrra suurem hüpertensiooni skoor, on 2,08 (usaldusvahemikuga: (1,684;2,571)) korda suurem šans saada hüpertensiooni diagnoos.

Järgnevalt vaadeldakse, millised näevad välja geeniandmeteta ja geeniandmetega mudelile joonistatud ROC-kõverad.



**Joonis 2.2.3.** Mudelitele 2.2.5 ja 2.2.6 joonistatud ROC-kõverad.

Jooniselt on näha, et kõverad langevad kokku ning geeniandmete mõju hüpertensiooni prognoosimisel pole eristatav, kuna vanuse ja kehamassiindeksi mõju on niivõrd tugev.

Järgnevalt jagati doonorid vanuse järgi kahte gruppi: 60aastased ja nooremad geenidoonorid ja üle 60aastased geenidoonorid ning koostati geeniandmeteta ja

geeniandmetega mudelid hüpertooniatõve diagnoosi esinemise prognoosimiseks üle 60aastastele ning 60aastastele ja noorematele doonoritele. Tulemused 60aastastele ja noorematele doonoritele on toodud tabelis 2.2.7 ja 2.2.8.

**Mudel 2.2.7.** Hüpertensiooni diagnoosi esinemine 60aastastel ja noorematel geenidoonoritel sõltuvalt soost ja kehamassiindeksist. Geeniandmeteta mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-6,43	0,238		<0,001
Naine	-0,73	0,086	0,48	<0,001
BMI	0,19	0,008	1,2	<0,001

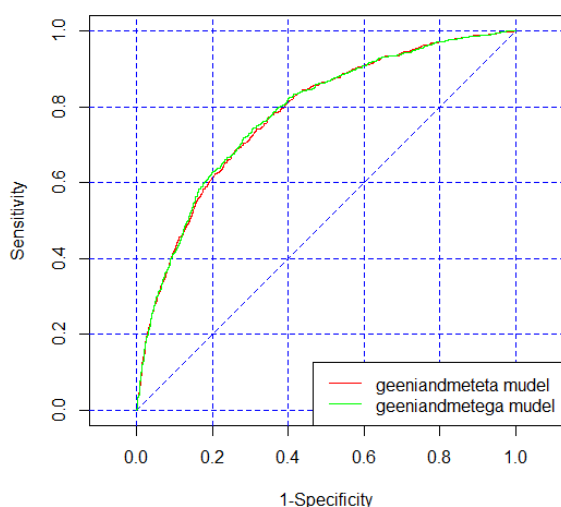
Mudelist selgub, et 60aastastel ja noorematel naisgeenidoonoritel on võrreldes meesdoonoritega 52% väiksem šanss saada hüpertooniatõve diagnoos (usaldusvahemikuga: (0,441;0,524)) ning 1 ühiku võrra suurema kehamassiindeksiga 60aastastel ja noorematel patsientidel on 1,2 korda suurem šanss saada hüpertensiooni diagnoos (usaldusvahemikuga: (1,199;1,219)). Vaadeldi ka geeniandmetega mudelit.

**Mudel 2.2.8.** Hüpertensiooni diagnoosi esinemine 60aastastel ja noorematel geenidoonoritel sõltuvalt soost, kehamassiindeksist ning hüpertensiooni skoori väärtusest. Geeniandmetega mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-7,14	0,33		<0,001
Naine	-0,73	0,086	0,48	<0,001
BMI	0,19	0,008	1,2	<0,001
Hüpertensiooni skoor	0,67	0,204	1,96	<0,001

60aastastel ja noorematel naistel võrreldes sama vanusegruppi kuuluvate meestega on 52% väiksem šanss (usaldusvahemikuga: (0,441;0,524)) hüpertensiooni diagnoosi esinemiseks, kehamassiindeksi kasvamine 1 ühiku võrra suurendab hüpertooniatõve diagnoosi saamist 60aastastel ja noorematel naistel 1,2 korda (usaldusvahemikuga: (1,199;1,219)) ning hüpertensiooni skoori kasvamisel 60aastastel ja noorematel 1 ühiku võrra suureneb šanss 1,96 korda (usaldusvahemikuga: (1,597;2,401)).

Joonistati 60aastastele ja noorematele patsientidele geeniandmeteta ja geeniandmetega mudelite põhjal ROC-kõverad, mis on toodud joonisel 2.2.4.

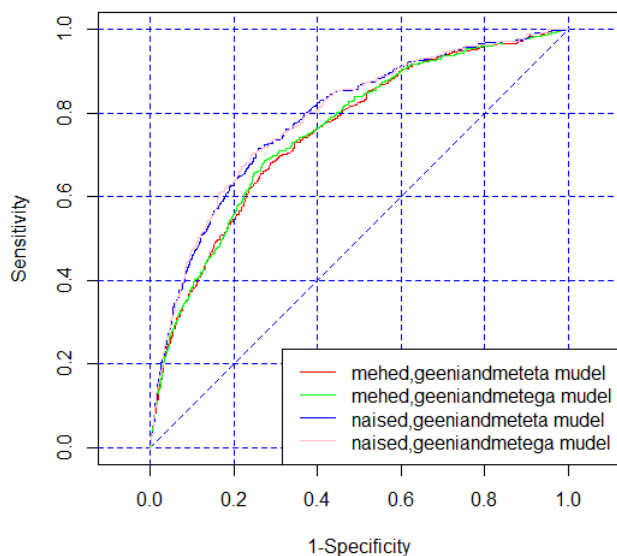


**Joonis 2.2.4.** Mudelitele 2.2.7 ja 2.2.8 joonistatud ROC-kõverad.

Geeniandmeteta mudelile joonistatud kõveraalune pindala on 0,781 ning geeniandmetega mudelile joonistatud kõveraalune pindala on 0,784. Ka siin langevad kõverad kokku ning võib öelda, et geeniandmete mõju pole eristatav, mis võib olla põhjustatud sellest, et soo ja kehamassiindeksi mõju on liialt tugev.

Järgnevalt jagati 60aastaste ja nooremate geenidoonorite grupp omakorda veel sooliselt kaheks ning joonistati neile ROC-kõverad, et näha, kas meeste ja naiste vahel antud vanusegrupis on erinevusi hüpertooniatõve diagnoosi esinemise prognoosimisel kasutades geeniandmetega või geeniandmeteta mudelit. Kõverad on toodud joonisel 2.2.5.

Jooniselt 2.2.5 on näha, et 60aastaste ja nooremate geenidoonorite seas on meestel ja naistel selge erinevus hüpertensiooni diagnoosi prognoosimise seisukohalt. Geeniandmete kasutamine siiski antud vanusegrupis suurt erinevust kaasa ei too. 60aastaste ja noematele naistele joonistatud geeniandmeteta mudel kõveraalune pindala on 0,788, meestel on sarnase mudeli puhul AUC võrdne 0,754. Kui lisada geeniandmed mudelisse, siis 60aastastele ja noematele naistele koostatud mudeli põhjal joonistatud ROC-kõveraalune pindala 0,789 ning meestele analoogse mudeli korral on AUC võrdne 0,758. Seega mõlema mudeli korral on geeniandmetega mudel pisut parem ning meeste puhul on geeniandmetega ja geeniandmeteta mudelite erinevus suurem. Antud juhul omab suurt mõju kehamassiindeks, mis ilmselt varjab geeniandmete mõju hüpertooniatõve diagnoosi prognoosimisel.



**Joonis 2.2.5.** 60aastastele ja noorematele meestele ja naistele geeniandmetega ja geeniandmeteta mudelitele joonistatud ROC-kõverad.

Vaadeldi ka vanemaid kui 60aastaseid. Nendele koostatud geeniandmeteta ja geeniandmetega mudelid on toodud tabelis 2.2.9 ja 2.2.10.

**Mudel 2.2.9.** Hüpertensiooni diagnoosi esinemine üle 60aastastel sõltuvalt soost ja kehamassiindeksist. Geeniandmeteta mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-2,39	0,243		<0,001
Naine	0,21	0,08	1,23	<0,001
BMI	0,1	0,009	1,11	<0,001

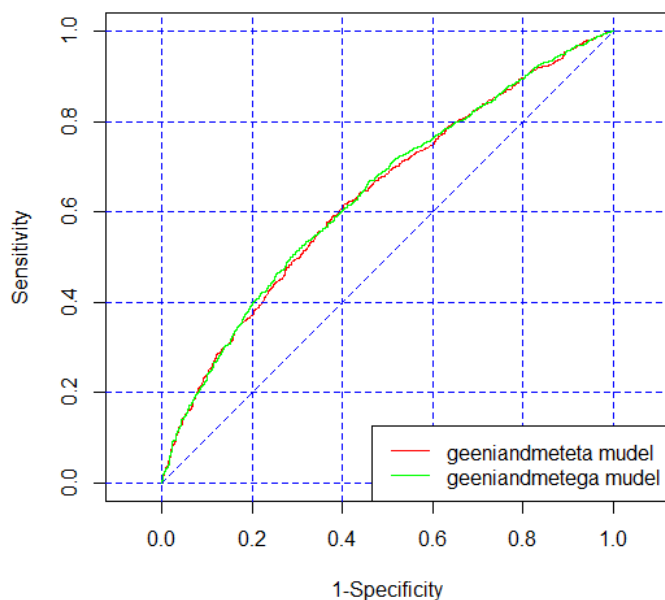
Üle 60aastastel naisgeenidoonoritel on 1,23 korda suurem šanss saada hüpertooniatõve diagnoos (usaldusvahemikuga: (1,136;1,335)) võrreldes samasse vanusegrupi kuuluvate meesdoonoritega ning 1 ühiku võrra kõrgema kehaindeksiga üle 60aastastel geenidoonoritel on 1,11 korda suurem šanss saada hüpertensiooni diagnoos (usaldusvahemikuga: (1,095;1,115)). Koostati analoogne geeniandmetega mudel üle 60aastastele geenidoonoritele.

**Mudel 2.2.10.** Hüpertensiooni diagnoosi esinemine üle 60aastastel sõltuvalt soost, kehamassiindeksist ning hüpertensiooni skoori väärtusest. Geeniandmetega mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-2,97	0,33		<0,001
Naine	0,2	0,204	1,22	0,01
BMI	0,1	0,009	1,11	<0,001
Hüpertensiooni skoor	0,51	0,195	1,67	0,008

Üle 60aastastel naistel on 1,22 korda suurem šanss (usaldusvahemikuga: (1,131;1,329)) saada hüpertooniatõve diagnoos võrreldes sama vanusegrupi meestega. Kehamassiindeksi kasvamisel ühe ühiku võrra suureneb üle 60aastastel geenidoonoritel šanss saada hüpertensiooni diagnoos 1,11 korda (usaldusvahemikuga (1,097;1,116)). Üle 60aastastel geenidoonoritel hüpertensiooniskoori arvutamisel suureneb šanss saada hüpertensiooni diagnoos 1,67 korda (usaldusvahemikuga: (1,377;2,034)).

Vaadeldi, kas geeniandmete mõju on hüpertooniatõve prognoosimisel märgatav vanusegrupis üle 60aastased geenidoonorid, selle tarvis joonistati mudelitele 2.2.10 ja 2.2.11 ROC-kõverad, mis on toodud joonisel 2.2.6.

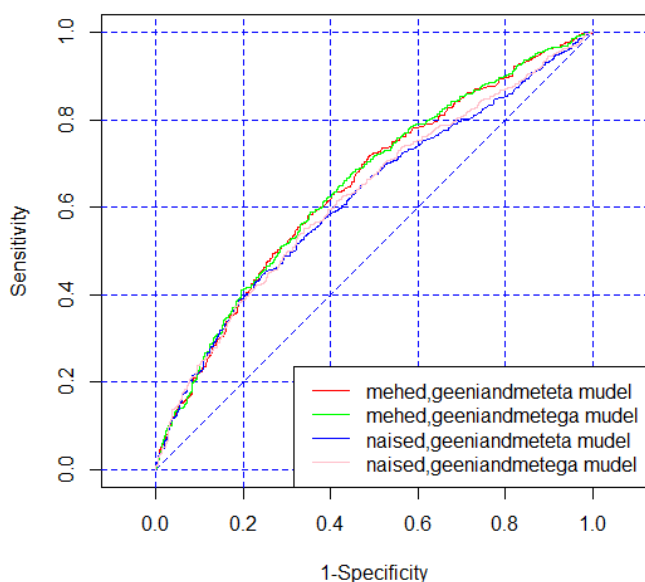


**Joonis 2.2.6.** Mudelitele 2.2.9 ja 2.2.10 joonistatud ROC-kõverad.

Geeniandmeteta mudelile joonistatud kõveraalne pindala on 0,637 ning geeniandmetega mudelile joonistatud kõveraalne pindala on 0,642. Kõverad langevad praktiliselt kokku ning vanusegrupis vanemad, kui 60aastased geenidoonorid pole võimalik eristada geeniandmete mõju, sest soo ning kehamassimõju on niivõrd suur.

Üle 60aastased geenidoonorid jaotati samuti sooliselt kahte gruppi, et näha, kas on meeste ja naiste vahel erinevus geeniandmetega ja geeniandmeteta mudelil hüpertooniatõve diagnoosi prognoosimise seisukohast. Joonistatud ROC-kõverad on toodud joonisel 2.2.7.

Järgnevalt jooniselt 2.2.7 selgub, et üle 60aastaste meeste seas geeniandmetega ja geeniandmeteta mudeli kasutamisel suurt erinevust pole, kuid naiste puhul on erinevus rohkem näha. Kui üle 60aastastele meestele koostatud geeniandmeteta mudelile joonistatud ROC-kõveraalne pindala on 0,649 ning naistel oli analoogse mudeli korras AUC võrdne 0,622, lisades mudelisse geeniandmed saadi meeste puhul kõveraalne pindala 0,652 ning naiste puhul sama näitaja 0,629. Seega geeniandmetel on mõju hüpertooniatõve diagnoosi prognoosimise seisukohalt ning see mõju on suurem, kui vaadelda üle 60aastaseid naisi eraldi grupina.



**Joonis 2.2.7.** Üle 60aastastele meestele ja naistele geeniandmetega ja geeniandmeteta mudelitele joonistatud ROC-kõverad.

### 2.3 Lineaarsete mudelite koostamine vererõhu tasemele

Kasutades lineaarset regressiooni koostati mudelid süstoolse ja diastoolse vererõhu taseme määramiseks, kui mudelisse on lisatud geenianndmed.

Esiteks uuriti, kuidas sõltub süstoolne vererõhk soost, vanuses, kehamassiindeksist ning arvutatud süstoolse vererõhu skoorist. Hinnangud on toodud tabelis 2.3.1 ja 2.3.2, kus eraldi mudelid koostati neile, kellel hüpertensiooni diagnoos esines ja neile, kellel mitte. Mõlemal juhul on tegemist statistiliselt oluliste mudelitega.

**Mudel 2.3.1.** Süstoolse vererõhu tase sõltuvalt soost, vanusest, kehamassiindeksist ja süstoolsest skoorist geenidoonoritel, kellel puudus hüpertensiooni diagnoos.

Parameeter	Hinnang	Standardviga	Olulisustõenäosus
Vabaliige	90,7	1,439	<0,001
Naine	-5,67	0,387	<0,001
Vanus	0,74	0,043	<0,001
BMI	0,3	0,01	<0,001
Süstoolne skoor	0,52	0,107	<0,001

Mudelist võib välja lugeda, et hüpertensiooni diagnoosita naiseks olemine vähendab süstoolse vererõhu taset 5,67 mmHg võrra, iga aasta lisandumisel kasvab süstoolse vererõhu tase 0,74 mmHg võrra, kehamassiindeksi suurenemisel 1 ühiku võrra kasvab süstoolse vererõhu tase 0,3 mmHg võrra ning süstoolse skoori taseme kasvamisel 1 ühiku võrra kasvab süstoolse vererõhu tase 0,52 mmHg võrra.

Järgnevalt tuuakse välja analoogne mudel geenidoonoritele, kellel esines hüpertensiooni diagnoos.

**Mudel 2.3.2.** Süstoolse vererõhu tase sõltuvalt vanusest, kehamassiindeksist ja süstoolsest skoorist geenidoonoritel, kellel esines hüpertensiooni diagnoos.

Parameeter	Hinnang	Standardviga	Olulisustõenäosus
Vabaliige	119,32	3,075	<0,001
Vanus	0,4	0,061	<0,001
BMI	0,13	0,023	<0,001
Süstoolne skoor	0,37	0,188	0,0475

Hüpertensiooni diagnoosi saanud geenidoonorite kõrgeneb vererõhk 0,4 mmHg võrra iga aasta lisandumisel, kehamassiindeksi suurenemisel 1 ühiku võrra kasvab süstoolse

vererõhu tase 0,13 mmHg ning süstoolse skoori suurenemisel 1 ühiku võrra kasvab süstoolse vererõhu tase 0,37 ühiku võrra. Sugu ei osutunud oluliseks süstoolse vererõhu taseme määramisel geenidonoritel, kellel diagnoos esines.

On näha, et süstoolsele vererõhule arvatud skoor mõjutab rohkem terveid patsiente, mis on ilmselt tingitud sellest, et enamus diagnoosiga geenidoonoreid saab rohtu ning see viib vererõhu normaalsele tasemele.

Koostati ka lineaarne mudel diastoolse vererõhu taseme määramiseks sõltuvalt soost, vanusest, kehamassiindeksist ning hüpertensiooni skoorist hüpertooniatõve diagnoosi esinemise lõikes. Tulemused on toodud tabelites 2.3.3 ja 2.3.4.

**Mudel 2.3.3.** Diastoolse vererõhu tase sõltuvalt soost, vanusest, kehamassiindeksist ja diastoolsest skoorist geenidonoritel, kellel puudus hüpertensiooni diagnoos.

Parameeter	Hinnang	Standardviga	Olulisustõenäosus
Vabaliige	56,38	0,947	<0,001
Naine	-2,68	0,261	<0,001
Vanus	0,58	0,029	<0,001
BMI	0,11	0,007	<0,001
Diastoolne skoor	0,47	0,115	<0,001

Naiseks olemine vähendab diastoolse vererõhu taset 2,68 mmHg võrra, iga aasta lisandumine tõstab diastoolset vererõhku 0,58 mmHg võrra, kehamassiindeksi suurenemisel 1 ühiku võrra suureneb diastoolse vererõhu tase 0,11 mmHg võrra ning diastoolse skoori suurenemisel ühe ühiku võrra suureneb diastoolse vererõhu tase 0,47 mmHg võrra.

Järgnevalt tuuakse mudel diastoolse vererõhu taseme määramiseks ka geenidonoritele, kellel esines hüpertensiooni diagnoos.

**Mudel 2.3.4.** Diastoolse vererõhu tase sõltuvalt soost, vanusest ja kehamassiindeksist geenidonoritel, kellel esines hüpertensiooni diagnoos.

Parameeter	Hinnang	Standardviga	Olulisustõenäosus
Vabaliige	83,1	1,589	<0,001
Naine	-1,33	0,438	0,002
BMI	0,3	0,038	<0,001
Vanus	-0,1	0,015	<0,001

Naiseks olemine vähendab diastoolse vererõhu taset 1,33 mmHg võrra ning kehamassiindeksi suurenemine ühe ühiku võrra tõstab diastoolse vererõhu taset 0,3 mmHg võrra ja aasta lisandumine vanusele vähendab diastoolse vererõhu taset 0,1 mmHg võrra.

Vaadeldes hüpertensiooni diagnoosiga geenidoonoreid selgus, et diastoolse skoori tase ei osutunud statistiliselt oluliseks, mis võib olla tingitud asjaolust, et doonorid, kellel diagnoos esines, tarvitavad vastavaid ravimeid, mis viivad diastoolse vererõhu taseme alla. Seega päriliku eelsoodumuse mõju ei pääse esile.

## 2.4 Alamandmestiku koostamine ja tulemused

Koostati alamandmestik, kuhu kuulusid kõik geenidoonorid, kes olid 30-60aastased ning kellede kehamassiindeks oli 25-30. Alamandmestikku kuulus 1150 geenidoonorit, kelledest hüpertooniatõve diagnoosi oli saanud 253 patsienti. Diagnoosi saanud patsientidele ehk juhtudele leiti juurde kontrollid, kes olid samas vanuses, samast soost ning kelle kehamassiindeks oli lähedane juhtude omale. Saadud andmestikus oli 241 juhtu ja 241 kontrolli, sest ülejäänud juhtudele ei olnud võimalik leida sobivaid kontrolle. Alamandmestik koostati, et elimineerida soo, vanuse ning kehamassimõju, et vaadelda, geenianndmete mõju hüpertensiooni diagnoosimisel. Koostati kaks logistilise regressiooni mudelit, millest ühes kasutati geenianndmeid ja teises mitte. Tulemused on toodud tabelites 2.4.1 ja 2.4.2.

**Mudel 2.4.1.** Hüpertensiooni diagnoosi esinemine 30-60aastastel patsientidel, kelle kehamassiindeks on 25-30. Geenianndmeteta mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-3,95	1,762		0,0249
BMI	0,14	0,064	1,15	0,0247

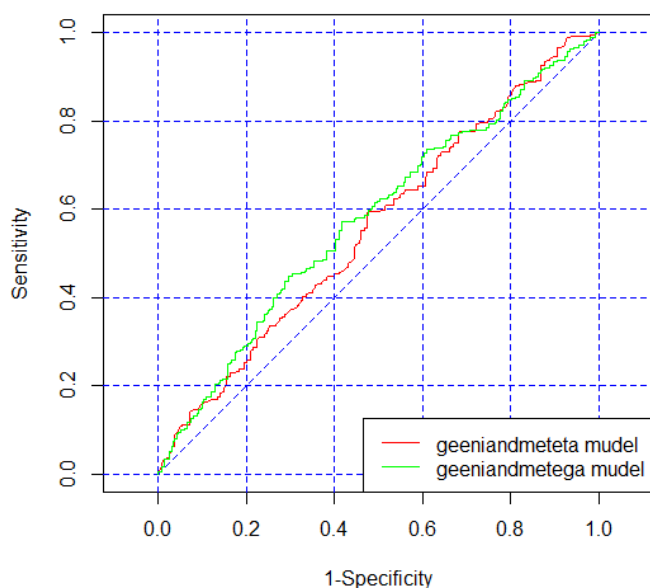
Geenianndmeteta mudelis osutus statistiliselt oluliseks geenidoonorite kehamassiindeks. Ühe ühiku võrra kõrgema kehamassiindeksiga patsientidel on 1,15 korda suurem risk saada hüpertensiooni diagnoos (usaldusvahemikuga: (1,083;1,231)) võrreldes 1 ühiku võrra madalama kehamassiindeksiga patsientidega. Järgmiseks koostati mudel, kus võeti arvesse ka geenianndmed, tulemused on toodud tabeli 2.4.2.

**Mudel 2.4.2.** Hüpertensiooni diagnoosi esinemine 30-60aastastel patsientidel, kelle kehamassiindeks on 25-30. Geeniandmetega mudel.

Parameeter	Hinnang	Standardviga	OR	Olulisustõenäosus
Vabaliige	-4,86	1,838		0,008
BMI	0,14	0,064	1,16	0,0243
Hüpertensiooni skoor	0,83	0,44	2,3	0,058

Statistiliselt olulisteks tunnusteks hüpertooniatõve diagnoosimisel osutusid kehamassiindeks ja hüpertensiooni skoor, mille autor otsustas lugeda samuti oluliseks, sest kuigi olulisustõenäosus oli veidi suurem kui 0,05 on see tingitud kehamassiindeksi mõjust. Ühe ühiku võrra kõrgema kehamassiindeksiga geenidoonoritel on 1,16 korda suurem šans saada hüpertensioonidiagnoos (usaldusvahemikuga: (1,084;1,233)) ning 1 ühiku võrra kõrgema hüpertensiooniskooriga geenidoonoritel on 2,3 korda suurem šans saada hüpertooniatõve diagnoos (usaldusvahemikuga: (1,481;3,568)).

Joonistatakse mudelitele 2.4.1 ja 2.4.2 vastav ROC-kõver, mis on toodud joonisel 2.4.1.



**Joonis 2.4.1.** Mudelitele 2.4.1 ja 2.4.2 joonistatud ROC-kõverad.

Geeniandmeteta mudeli kõveraalune pindala on 0,559 ning geeniandmetega mudeli kõveraalune pindala on 0,58. On näha, et geeniandmetega mudel annab lisainformatsiooni hüpertooniatõve diagnoosimiseks, kui on eemaldatud soo ja vanuse mõju.

## 2.5 Ümberklassifitseerimise indeksid

Antud peatükis arvutatakse ümberklassifitseerimise indeksid, et näha, kas geenianndmetega mudelid osutuvad paremaks hüpertooniatõve diagnoosi prognoosimisel. Ümberklassifitseerimise indekse arutamiseks osutuvad vajalikuks ümberklassifitseerimise tabelid, mis koostatakse eraldi diagnoosiga ja diagnoosita geenidoonoritele. Esiteks koostati ümberklassifitseerimise indeksi arutamiseks vajalikud tabelid naispatsientide tarvis, tulemused on toodud tabelites 2.5.1 ja 2.5.2. Tabelite koostamisel on jäetud välja need geenidoonorid, kellede puhul oli puuduvaid andmeid. Ümberklassifitseerimise tabelite koostamisel on otsustuspiiridega jaotud patsiendid 3 gruppi: madala [0,0.5), keskmise [0.5,0.75) ja kõrge geneetilise riskiga [0.75,1] geenidoonoriteks. Ning kõikidel juhtudel arvutati välja pidev NRI ehk  $NRI(>0)$ .

**Tabel 2.5.1.** Hüpertensiooni diagnoosiga patsientidele koostatud reklassifitseerimise tabel.

		Geenianndmetega mudel		
		[0,0.5)	[0.5,0.75)	[0.75,1]
Geenianndmeteta mudel	[0,0.5)	886	66	0
	[0.5,0.75)	47	992	49
	[0.75,1]	0	40	474

**Tabel 2.5.2.** Hüpertensiooni diagnoosita patsientidele koostatud reklassifitseerimise tabel.

		Geenianndmetega mudel		
		[0,0.5)	[0.5,0.75)	[0.75,1]
Geenianndmeteta mudel	[0,0.5)	4083	51	0
	[0.5,0.75)	40	569	16
	[0.75,1]	0	14	138

Prognoosikategooria paranes 115 diagnoosiga ja 54 diagnoosita geenidoonoril ning halvenes 87 diagnoosiga ja 67 diagnoosita geenidoonoril- kokku muutus korrektsemaks 28 hüpertensiooni diagnoosiga patsiendil. Arvutati välja ka  $NRI(>0)$ , mille väärtuseks tuli 0,123 usaldusvahemikuga (0,075; 0,171) ning olulisustõenäosuseväärtuseks saadi  $p=0$ , mis tähendab, et geenianndmetega mudel erineb geenianndmeteta mudelist ning kuna  $NRI(>0)$  on positiivne, siis see näitab, et geenianndmeid kasutades saadakse parem mudel hüpertooniatõve diagnoosimiseks. Vaadatakse ümberklassifitseerimist ka riskigruppide lõikes, tulemused on toodud tabelis 2.5.3.

**Tabel 2.5.3.** Ümberklassifitseerimine riskiskooride lõikes kõikide geenidoonorite hulgas.

Geeniandmeteta mudel	Haigete arv	Tervete arv	%, kes klassifitseeriti üles,haiged	%, kes klassifitseeriti üles,terved	%, kes klassifitseeriti alla,haiged	%, kes klassifitseeriti alla,terved
[0,0.5)	952	4134	6,90%	1,20%		
[0.5,0.75)	1088	625	4,50%	2,60%	4,30%	6,40%
[0.75,1]	514	152			7,80%	9,20%

Tabelist 2.5.3 on näha, et madala riskiskooriga grupis on protsentuaalselt ligikaudu 5 korda rohkem geenidoonoreid klassifitseeritud õiges suunas, kasutades geeniandmeid. Keskmise riskiskooriga grupis on sarnane tendents: õiges suunas klassifitseeritud on geeniandmete kasutamisel protsentuaalselt rohkem. Kõrgema riskiskooriga grupis on samuti õiges suunas klassifitseeritud geenidoonoreid rohkem, kuid protsentuaalne erinevus pole niivõrd suur kui madala ja keskmise riskiskooriga gruppides. Personaalse meditsiini seisukohalt ongi tähtis keskmise riskiskooriga patsiendid, sest sellisel juhul on võimalus, et geeniandmed annavad uut lisainformatsiooni, et prognoosida antud kontekstis hüpertensiooni esinemist.

Koostati ümberklassifitseerimise tabelid meestele, mis on toodud tabelites 2.5.4 ja 2.5.5.

**Tabel 2.5.4.** Hüpertensiooni diagnoosiga meespatsientidele koostatud reklassifitseerimise tabel.

		Geeniandmetega mudel meestele		
		[0,0.5)	[0.5,0.75)	[0.75,1]
Geeniandmeteta mudel meestele	[0,0.5)	507	29	0
	[0.5,0.75)	24	469	15
	[0.75,1]	0	16	144

**Tabel 2.5.5.** Hüpertensiooni diagnoosita meespatsientidele koostatud reklassifitseerimise tabel.

		Geeniandmetega mudel meestele		
		[0,0.5)	[0.5,0.75)	[0.75,1]
Geeniandmeteta mudel meestele	[0,0.5)	1839	23	0
	[0.5,0.75)	20	292	4
	[0.75,1]	0	5	45

Tabelitest 2.5.4 ja 2.5.5 selgub, et prognoosikategooria paranes 44 hüpertooniatõve diagnoosiga ja 25 diagnoosita meessoost geenidoonoril ning halvenes 40 diagnoosiga patsiendil ning 27 diagnoosita meessoost geenidoonoril. Kokku muutust prognoosikategooria korrektsemaks 4 hüpertooniatõve diagnoosiga patsiendil. NRI(>0) väärtuseks saadi 0,108 usaldusvahemikuga (0,038;0,178) ning olulisustõenäosus oli 0,0025, mis tähendab, et geenianndmetega mudel erineb geenianndmeteta mudelist ning kuna NRI(>0) on positiivne, siis geenianndmeid kasutades saadakse parem mudel prognoosimaks hüpertensiooni diagnoosi esinemist. Vaadeldakse meessoost patsientide jagunemist riskigruppide lõikes, tulemused on toodud tabelis 2.5.6.

**Tabel 2.5.6.** Ümberklassifitseerimine riskiskooride lõikes meessoost geenidoonorite hulgas.

<b>Geeniandmeteta mudel</b>	<b>Haigete arv</b>	<b>Tervete arv</b>	<b>%, kes klassifitseeriti üles,haiged</b>	<b>%, kes klassifitseeriti üles,terved</b>	<b>%, kes klassifitseeriti alla,haiged</b>	<b>%, kes klassifitseeriti alla,terved</b>
[0,0.5)	536	1862	5,4%	1,2%		
[0.5,0.75)	508	316	3,0%	1,3%	4,7%	6,3%
[0.75,1]	160	50			10,0%	10,0%

Ülaltoodud tabelist on näha, et meessoost geenidoonoritel, kes kuuluvad madala riskiskooriga gruppi, on õiges suunas klassifitseeritud protsentuaalselt 4 korda rohkem, kasutades geenianndmeid. Ka keskmise riskiskooriga grupis on õiges suunas klassifitseeritud meessoost geenidoonoreid protsentuaalselt rohkem. Kõrge riskiskooriga gruppi kuuluvatele meesgeenidoonoritel geenianndmete kasutamine lisainformatsiooni ei anna: õiges ja vales suunas klassifitseeritud on protsentuaalselt võrdselt.

Arvutatakse välja reklassifitseerimise tabelid naissoost patsientide tarvis, tulemused on toodud tabelites 2.5.7 ja 2.5.8.

**Tabel 2.5.7.** Hüpertensiooni diagnoosiga naispatsientidele koostatud reklassifitseerimise tabel.

		Geeniandmetega mudel naistele		
		[0,0.5)	[0.5,0.75)	[0.75,1]
Geeniandmeteta mudel naistele	[0,0.5)	401	27	0
	[0.5,0.75)	17	494	28
	[0.75,1]	0	27	356

**Tabel 2.5.8.** Hüpertensiooni diagnoosita naispatsientidele koostatud reklassifitseerimise tabel.

		Geeniandmetega mudel naistele		
		[0,0.5)	[0.5,0.75)	[0.75,1]
Geeniandmeteta mudel naistele	[0,0.5)	2260	20	0
	[0.5,0.75)	17	260	7
	[0.75,1]	0	9	110

Prognosikategooria paranes 55 hüpertensiooni diagnoosiga ja 26 diagnoosita naissoost geenidoonoril ning halvenes 44 diagnoosiga ja 27 diagnoosita naissoost geenidoonoril. Prognos muutus korrektsemas 11 diagnoosiga patsiendil. NRI(>0) väärtuseks saadi 0,138 usaldusvahemikuga (0,072;0,203) olulisustõenäosus oli <0,001, seega geeniandmetega mudel erineb geeniandmeteta mudelist ning geeniandmetega mudel on parem hüpertooniatõve diagnoosi esinemise prognoosimiseks, sest NRI(>0) on positiivne. Tabelis 2.5.8 tuakse välja naissoost geenidoonorite jagunemine riskigruppide lõikes.

**Tabel 2.5.8.** Ümberklassifitseerimine riskiskooride lõikes naissoost geenidoonorite hulgas.

Geeniandmeteta mudel	Haigete arv	Tervete arv	%, kes klassifitseeriti üles,haiged	%, kes klassifitseeriti üles,terved	%, kes klassifitseeriti alla,haiged	%, kes klassifitseeriti alla,terved
[0,0.5)	428	2280	6,3%	0,9%		
[0.5,0.75)	539	284	5,2%	2,5%	3,2%	6,0%
[0.75,1]	383	119			7,0%	7,6%

Tabelist 2.5.8 on näha, et madala riskiskooriga naissoost geenidoonoreid, kes on õiges suunas klassifitseeritud, on protsentuaalselt 7 korda rohkem, kasutades geeniandmeid. Keskmise riskiskooriga geenidoonorite hulgas on õigesti klassifitseeritud ligikaudu 2 korda rohkem, see tähendab, et ka selles grupis annavad geeniandmed vajalikku uut informatsiooni. Kõrgemas riskigrupis on protsentuaalne erinevus õigesti ja valesti klassifitseeritute hulgas väiksem, kuid õiges suunas klassifitseeritud on siiski rohkem.

Ümberklassifitseerimise tabelid koostati 60aastastele ja noorematele geenidoonoritele, tulemused on toodud tabelites 2.5.9 ja 2.5.10.

**Tabel 2.5.9.** Hüpertensiooni diagnoosiga 60aastastele ja noorematele patsientidele koostatud reklassifitseerimise tabel.

		Geeniandmetega mudel ≤60aastased		
		[0,0.5)	[0.5,0.75)	[0.75,1]
Geeniandmeteta mudel ≤60aastased	[0,0.5)	663	7	0
	[0.5,0.75)	6	130	0
	[0.75,1]	0	2	27

**Tabel 2.5.10.** Hüpertensiooni diagnoosita 60aastastele ja noorematele patsientidele koostatud reklassifitseerimise tabel.

		Geeniandmetega mudel ≤60aastased		
		[0,0.5)	[0.5,0.75)	[0.75,1]
Geeniandmeteta mudel ≤60aastased	[0,0.5)	3708	15	0
	[0.5,0.75)	7	83	2
	[0.75,1]	0	0	10

Prognoosikategooria paranes 7 hüpertooniatõve diagnoosiga ja 7 diagnoosita 60aastasel ja nooremal patsiendil ning halvenes 8 diagnoosiga ja 17 diagnoosita patsiendil. NRI(>0) väärtuseks saadi 0,123 usaldusvahemikuga (0,048; 0,198), olulisustõenäosus oli 0,00127 ning seega geeniandmetega mudel erineb geeniandmeteta mudelist. Kuna NRI(>0) oli positiivne, siis kasutades geeniandmeid saadakse parem mudel prognoosimaks hüpertooniatõve diagnoosi esinemist. Tabelis 2.5.11 tuuakse välja 60aastaste ja nooremate geenidoonorite klassifitseerimine riskigruppide lõikes.

**Tabel 2.5.11.** Ümberklassifitseerimine riskiskooride lõikes 60aastastele ja nooremate geenidoonorite hulgas.

Geeniandmeteta mudel	Haigete arv	Tervete arv	%, kes klassifitseeriti üles,haiged	%, kes klassifitseeriti üles,terved	%, kes klassifitseeriti alla,haiged	%, kes klassifitseeriti alla,terved
[0,0.5)	670	3723	1,0%	0,4%		
[0.5,0.75)	136	92	0,0%	2,2%	4,4%	7,6%
[0.75,1]	29	10			6,9%	0,0%

Madalasse riskigrupi kuuluvatest 60aastastest ja noorematest geenidoonoritest klassifitseeriti õiges suunas protsentuaalselt 2 korda rohkem võrreldes vales suunas klassifitseeritutelega, kasutades geeniandmeid. Klassifitseerimine ei töötanud hästi keskmise ja kõrge riskiskooriga grupis, sest antud gruppides on võrreldes madala riskiskoori grupiga tunduvalt vähem geenidoonoreid, seega ei saa tõsiseid järeldusi teha õiges suunas klassifitseerimise kohta.

Ümberklassifitseerimise tabelid koostati üle 60aastastele patsientidele, et näha, kas on erinevusi geeniandmete kasutamisel. Tulemused on toodud tabelites 2.5.12 ja 2.5.13.

**Tabel 2.5.12.** Hüpertensiooni diagnoosiga üle 60aastastele patsientidele koostatud reklassifitseerimise tabel.

		Geeniandmetega mudel		
		>60aastased		
		[0,0.5)	[0.5,0.75)	[0.75,1]
Geeniandmeteta mudel >60aastased	[0,0.5)	187	32	0
	[0.5,0.75)	34	1165	31
	[0.75,1]	0	20	250

**Tabel 2.5.13.** Hüpertensiooni diagnoosita üle 60aastastele patsientidele koostatud reklassifitseerimise tabel.

		Geeniandmetega mudel >60aastased		
		[0,0.5)	[0.5,0.75)	[0.75,1]
Geeniandmeteta mudel >60aastased	[0,0.5)	210	34	0
	[0.5,0.75)	51	720	10
	[0.75,1]	0	7	54

Ümberklassifitseerimise tabelitest võib välja lugeda, et prognoosikategooria paranes 63 hüpertensiooni diagnoosiga ja 58 diagnoosita üle 60aastasel geenidoonoril ning halvenes 54 diagnoosiga ja 44 hüpertooniatõve diagnoosita geenidoonoril. Korrektsemaks muutus prognoosikategooria 9 diagnoosiga ja 14 diagnoosita üle 60aastasel geenidoonoril. NRI(>0) väärtuseks saadi 0,079 usaldusvahemikuga (0,003;0,155), olulisustõenäosus oli 0,04 ning seega geeniandmetega mudel erineb geeniandmeteta mudelist, kuna NRI(>0) oli positiivne, siis geeniandmeid kasutades saadakse parem mudel hüpertooniatõve prognoosimiseks. Vaadeldi klassifitseerimist üle 60aastaste geenidoonorite hulgas riskigruppide lõikes, tulemused on toodud tabelis 2.5.14.

**Tabel 2.5.14.** Ümberklassifitseerimine riskiskooride lõikes üle 60aastaste geenidoonorite hulgas.

Geeniandmeteta mudel	Haigete arv	Tervete arv	%, kes klassifitseeriti üles,haiged	%, kes klassifitseeriti üles,terved	%, kes klassifitseeriti alla,haiged	%, kes klassifitseeriti alla,terved
[0,0.5)	219	244	14,6%	13,9%		
[0.5,0.75)	1230	781	2,5%	1,3%	2,8%	6,5%
[0.75,1]	270	61			7,4%	11,5%

Tabelist selgub, et klassifitseerimine 60aastaste ja vanemate geenidoonorite jaoks töötab hästi kõigis kolmes riskiskoori grupis: õiges suunas klassifitseeritud on kõikides riskigruppides protsentuaalselt rohkem kui vales suunas klassifitseeritud. Madalas riskikrupis on protsentuaalne erinevus väike õigesti ja valesti klassifitseeritud geenidoonorite hulgas, kuid keskmises riskigrupis on see ligikaudu 2 kordne. See näitab, et geeniandmed annavad selle vanusegrupi jaoks olulist informatsiooni hüpertensiooni

prognoosimisel ning ka kõrges riskigrupis on erinevus märgatav: õigesti klassifitseeritud on protsentuaalselt enam, kui valesti klassifitseeritud.

Ümberklassifitseerimine töötab hästi üle 60aastaste geenidoonorite jaoks, sellepärast otsustati otsustuspiirid teha väiksemaks, et näha täpsemini geenidoonorite jagunemist. Antud otsus tugines ka asjaolul, et enamus haigeid kuuluvad sellesse vanusegruppi. Tulemused on toodud tabelites 2.5.15 ja 2.5.16. Otsustuspiirid jaotati kuueks grupiks: [0,0.5), [0.5,0.6), [0.6,0.7), [0.7,0.8), [0.8,0.9) ja [0.9,1].

**Tabel 2.5.15.** Hüpertensiooni diagnoosiga üle 60aastastele patsientidele koostatud reklassifitseerimise tabel kitsamate otsustuspiiridega.

		Geeniandmetega mudel >60aastased					
		[0,0.5)	[0.5,0.6)	[0.6,0.7)	[0.7,0.8)	[0.8,0.9)	[0.9,1]
Geeniandmeteta mudel >60aastased	[0,0.5)	187	32	0	0	0	0
	[0.5,0.6)	34	365	64	0	0	0
	[0.6,0.7)	0	66	437	48	0	0
	[0.7,0.8)	0	0	44	284	17	0
	[0.8,0.9)	0	0	0	7	102	6
	[0.9,1]	0	0	0	0	2	24

**Tabel 2.5.16.** Hüpertensiooni diagnoosita üle 60aastastele patsientidele koostatud reklassifitseerimise tabel kitsamate otsustuspiiridega.

		Geeniandmetega mudel >60aastased					
		[0,0.5)	[0.5,0.6)	[0.6,0.7)	[0.7,0.8)	[0.8,0.9)	[0.9,1]
Geeniandmeteta mudel >60aastased	[0,0.5)	210	34	0	0	0	0
	[0.5,0.6)	51	328	37	0	0	0
	[0.6,0.7)	0	30	240	25	0	0
	[0.7,0.8)	0	0	13	89	1	0
	[0.8,0.9)	0	0	0	5	20	0
	[0.9,1]	0	0	0	0	0	3

Ümberklassifitseerimise tabelist on näha, et prognoosikategooria paranes 167 hüpertensiooni diagnoosiga ning 99 hüpertooniatõve diagnoosita üle 60aastasel geenidoonoril ja halvenes 153 diagnoosiga ning 97 diagnoosita üle 60aastasel geenidoonoril. Seega prognoosikategooria muutus korrektsemaks 14 diagnoosiga ning 2

diagnoosita üle 60aastasel geenidoonoril. Kuna  $NRI(>0)=0,079$  väärtus on positiivne, siis võib öelda, et täpsema prognoosi saab kasutades geeniandmetega mudelit.

Vaadatakse eraldi veel, millised on ümberklassifitseerimise tulemused eraldi üle 60aastaste meeste ja naiste jaoks.

Esiteks tuuakse välja ümberklassifitseerimise tabelid üle 60aastaste naiste jaoks, tulemused on toodud tabelites 2.5.17 ja 2.5.18.

**Tabel 2.5.17.** Hüpertensiooni diagnoosiga üle 60aastastele naispatsientidele koostatud reklassifitseerimise tabel kitsamate otsustuspiiridega.

		Geeniandmetega mudel >60aastased naised					
		[0,0.5)	[0.5,0.6)	[0.6,0.7)	[0.7,0.8)	[0.8,0.9)	[0.9,1]
Geeniandmeteta mudel >60aastased naised	[0,0.5)	35	12	0	0	0	0
	[0.5,0.6)	18	189	47	0	0	0
	[0.6,0.7)	0	44	292	56	0	0
	[0.7,0.8)	0	0	37	193	15	0
	[0.8,0.9)	0	0	0	6	68	0
	[0.9,1]	0	0	0	0	0	15

**Tabel 2.5.18.** Hüpertensiooni diagnoosita üle 60aastastele naispatsientidele koostatud reklassifitseerimise tabel kitsamate otsustuspiiridega.

		Geeniandmetega mudel >60aastased naised					
		[0,0.5)	[0.5,0.6)	[0.6,0.7)	[0.7,0.8)	[0.8,0.9)	[0.9,1]
Geeniandmeteta mudel >60aastased naised	[0,0.5)	32	8	0	0	0	0
	[0.5,0.6)	20	174	26	0	0	0
	[0.6,0.7)	0	29	176	15	0	0
	[0.7,0.8)	0	0	7	63	3	0
	[0.8,0.9)	0	0	0	3	11	1
	[0.9,1]	0	0	0	0	0	1

Prognoosikategooria paranes 130 hüpertooniatõve diagnoosiga ja 59 diagnoosita naissoost üle 60aastasel patsiendil ning halvenes 105 diagnoosiga ja 53 diagnoosita naissoost üle 60aastasel geenidoonoril. Prognoosikategooria muutus korrektsemaks 25 hüpertooniatõve diagnoosiga patsiendil ning halvenes 6 diagnoosita üle 60aastasel naissoost geenidoonoril.  $NRI(>0)$  väärtuseks oli 0,135 (usaldusvahemikuga: (0,033;0,238)) olulisustõenäosusega

0,0096, seega geeniaandmetega mudel osutub paremaks hüpertooniatõve prognoosimiseks üle 60aastaste naisgeenidoonorite seas. Analoogsed tabelid koostati ka üle 60aastaste meessoost patsientide jaoks, tulemused on toodud tabelites 2.5.19 ja 2.5.20.

**Tabel 2.5.19.** Hüpertensiooni diagnoosiga üle 60aastastele meespatsientidele koostatud reklassifitseerimise tabel kitsamate otsustuspiiridega.

		Geeniaandmetega mudel >60aastased mehed					
		[0,0.5)	[0.5,0.6)	[0.6,0.7)	[0.7,0.8)	[0.8,0.9)	[0.9,1]
Geeniaandmeteta mudel >60aastased mehed	[0,0.5)	144	13	0	0	0	0
	[0.5,0.6)	12	164	15	0	0	0
	[0.6,0.7)	0	22	175	9	0	0
	[0.7,0.8)	0	0	8	87	3	0
	[0.8,0.9)	0	0	0	4	24	1
	[0.9,1]	0	0	0	0	1	10

**Tabel 2.5.20.** Hüpertensiooni diagnoosita üle 60aastastele meespatsientidele koostatud reklassifitseerimise tabel kitsamate otsustuspiiridega.

		Geeniaandmetega mudel >60aastased mehed					
		[0,0.5)	[0.5,0.6)	[0.6,0.7)	[0.7,0.8)	[0.8,0.9)	[0.9,1]
Geeniaandmeteta mudel >60aastased mehed	[0,0.5)	207	8	0	0	0	0
	[0.5,0.6)	12	136	13	0	0	0
	[0.6,0.7)	0	11	82	5	0	0
	[0.7,0.8)	0	0	3	31	1	0
	[0.8,0.9)	0	0	0	2	3	1
	[0.9,1]	0	0	0	0	0	2

Ümberklassifitseerimise tabelist selgub, et prognoosikategooria paranes 41 diagnoosiga ja 28 hüpertooniatõve diagnoosita üle 60aastasel meessoost patsiendil ning halvenes 47 diagnoosiga ja 28 diagnoosita üle 60aastasel meespatsiendil. Seega prognoosikategooria ei paranenud mitte ühelgi üle 60aastasel meessoost patsiendil. NRI(>0) väärtuseks tuli 0,012 (usaldusvahemikuga: (-0,102;0,126)), olulisustõenäosusega 0,8426. Geeniaandmetega mudelit kasutades ei saada paremat mudelit prognoosimaks hüpertooniatõve diagnoosi esinemist üle 60aastaste meesgeenidoonorite seas.

Ümberklassifitseerimise tabelid koostati alamandmestikule, kus oli püütud elimineerida soo, vanuse ja kehamassiindeksi mõju. Saadud tulemused on toodud tabelites 2.5.21 ja 2.5.22.

**Tabel 2.5.21.** Hüpertensiooni diagnoosiga 30-60aastastele patsientidele, kellede kehamassiindeks on 25-30 koostatud reklassifitseerimise tabel kitsamate otsustuspiiridega.

		Geeniandmetega mudel		
		[0,0.5)	[0.5,0.6)	[0.6,0.7)
Geeniandmeteta mudel	[0,0.5)	85	23	0
	[0.5,0.6)	16	98	19

**Tabel 2.5.22.** Hüpertensiooni diagnoosita 30-60aastastele patsientidele, kellede kehamassiindeks on 25-30 koostatud reklassifitseerimise tabel kitsamate otsustuspiiridega.

		Geeniandmetega mudel		
		[0,0.5)	[0.5,0.6)	[0.6,0.7)
Geeniandmeteta mudel	[0,0.5)	109	21	0
	[0.5,0.6)	24	77	10

Alamandmestikus, kus oli elimineeritud soo, vanuse ning kehamassimõju, muutus prognoosikategooria paremaks 42 hüpertensiooni diagnoosiga ja 24 diagnoosita patsiendil ning prognoosikategooria halvenes 16 diagnoosiga ja 31 diagnoosita patsiendil. Kokku muutus prognoosikategooria korrektsemaks 26 diagnoosiga geenidoonoril. NRI(>0) väärtuseks saadi 0,208 (usaldusvahemikuga: (0,029;0,386)), olulisustõenäosusega 0,023. Seega kasutades geeniandmeid saadaks parem mudel prognoosimaks hüpertooniatõve esinemist geenidoonorite seas, kus on sarnaste tunnustega patsiendid. Kas selle andmestiku pealt vaadati, kuidas toimib klassifitseerimine riskigruppide lõikes, tulemused on toodud tabelis 2.5.23.

**Tabel 2.5.23.** Ümberklassifitseerimine riskiskooride lõikes alamandmestikus.

Geeniandmeteta mudel	Haigete arv	Tervete arv	%, kes klassifitseeriti üles,haiged	%, kes klassifitseeriti üles,terved	%, kes klassifitseeriti alla,haiged	%, kes klassifitseeriti alla,terved
[0,0.5)	108	130	21,3%	16,2%		
[0.5,0.75)	133	111	14,3%	9,0%	12,0%	21,6%

Tabelist 2.5.23 on näha, et klassifitseerimine töötab hästi ka alamandmestiku jaoks: klassifitseeritud õiges suunas on protsentuaalselt rohkem mõlemas riskigrupis. On näha, et geenianndmed annavad uut informatsiooni hüpertensiooni prognoosimisel, kui muudest faktoritest tulenev mõju on viidud minimaalseks.

## 2.6 Kokkuvõtte erinevate mudelite prognoosivõimest

Vaadeldes kokkuvõtvalt erinevate mudelite prognoosivõimet on näha, et hüpertensiooni diagnoosi korral on geneetilisel skooril küll mõju, kuid see mõju on suhteliselt nõrk võrreldes muude teguritega. Kui vaadelda lineaarseid mudeleid, mis koostati süstoolse ja diastoolse vererõhu taseme määramiseks, siis on selge, et kõrgem geneetiline skoor toob kaasa vererõhu tõusu ning seda just geenidoonoritel, kellel hüpertensiooni diagnoos puudus, sest ilmselt tarvitavad diagnoosiga patsiendid vastavaid ravimeid, mis viivad vererõhu normaalsele tasemele.

Et vältida muude tegurite mõju moodustati andmestik, kuhu valiti 30-60aastased geenidoonorid, kellede kehamassiindeks oli 25-30. Igale juhule leiti vastav kontroll ning seejärel koostati mudelid hüpertooniatõve prognoosimiseks. Seega olles elimineerinud soo, vanuse ja kehamassiindeksi mõju, leiti, et geenidel on mõju hüpertensiooni diagnoosi esinemiseks.

Geenianndmeteta ja geenianndmetega mudelitele koostati ka ümberklassifitseerimise tabelid ning arvutati välja NRI.

Kokkuvõtlikult on käesoleva magistritöö tulemused toodud tabelis 2.6.1.

**Tabel 2.6.1.** Erinevate mudelite võrdlused.

	<b>Geeni- andmeteta AUC</b>	<b>Geeni- andmetega AUC</b>	<b>NRI(&gt;0)</b>	<b>CI</b>	<b>p</b>
<b>Kõik geenidoonorid</b>	0,84	0,842	0,123	(0,075;0,171)	0
<b>Mehed</b>	0,805	0,806	0,108	(0,038;0,178)	0,0025
<b>Naised</b>	0,873	0,874	0,138	(0,072;0,203)	<0,001
<b>Kõik 60aastased ja Nooremad</b>	0,781	0,784	0,123	(0,048;0,198)	0,00127

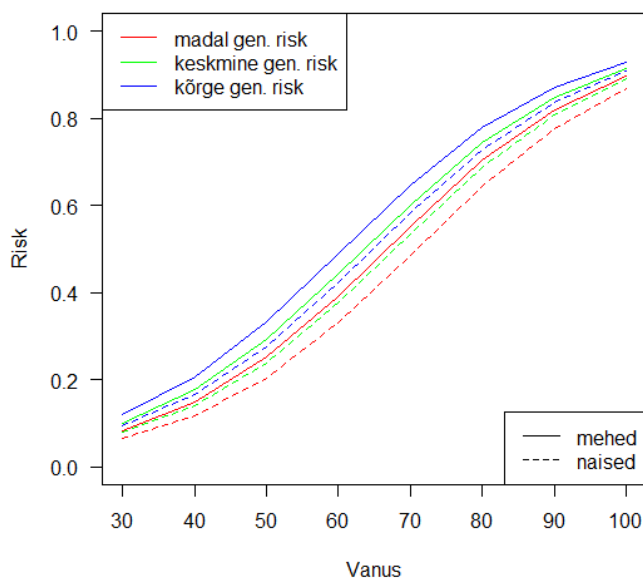
	<b>Geeni- andmeteta AUC</b>	<b>Geeni- andmetega AUC</b>	<b>NRI(&gt;0)</b>	<b>CI</b>	<b>p</b>
<b>60aastased ja nooremad mehed</b>	0,754	0,758	0,160	(0,062;0,259)	0,00147
<b>60aastased ja nooremad naised</b>	0,788	0,789	0,068	(0,049;0,185)	0,2572
<b>Kõik üle 60aastased</b>	0,637	0,642	0,079	(0,003;0,155)	0,04078
<b>Mehed üle 60aastased</b>	0,649	0,652	0,012	(-0,102;0,126)	0,84267
<b>Naised üle 60aastased</b>	0,622	0,628	0,135	(0,033;0,238)	0,0096
<b>30-60aastased geenidonorid kehamassiindegiga 25-30</b>	0,559	0,58	0,208	(0,0289;0,386)	0,02276

Kui võrrelda omavahel erinevate mudelite AUC-sid, siis on näha, et erinevates gruppides on geenianandmetega mudeli AUC suhteliselt lähedane geenianandmeteta mudeli AUC-le. Igas grupis on siiski näha, et geenianandmetega mudeli AUC on pisut kõrgem kui geenianandmeteta mudeli AUC. Suurem erinevus on koostatud alamandmestiku puhul, kus mudelite erinevus paistab rohkem silma. Kui võrrelda NRI-sid erinevate gruppide vahel on näha, et geenianandmetel on igas grupis mõju hüpertooniatõve diagnoosi esinemisele. See mõju pääseb kõige rohkem esile koostatud alamandmestiku puhul, kus oli üritatud elimineerida soo, vanuse ja kehamassi mõju, samuti on näha, et geenianandmete mõju hüpertooniatõve prognoosimiseks on tugevam üle 60aastaste naiste seas ning 60aastaste ja nooremate meeste seas. Seega võib öelda, et geenimarkerite prognoosivõime hüpertooniatõve ja sellega kaasnevalt ka südame-veresoonkonnahaiguste riski hindamisel on eristatav, kui elimineerida soo, vanuse ja kehamassiindeksi mõju. Sest antud juhul on väliste tegurite mõju niivõrd suur, et peidab geenimarkerite mõju ära.

## 2.7 Hüpertensiooni riski hindamine

Geneetilistesse riskigruppidesse jagamine käis sarnaselt eelmisele peatükile, kus madal geneetiline risk tähendas riskigrupi [0,0.5), keskmine geneetiline riskigrupp tähendab riskigrupi [0.5,0.75) ning kõrge geneetiline risk tähendab riskigrupi [0.75,1].

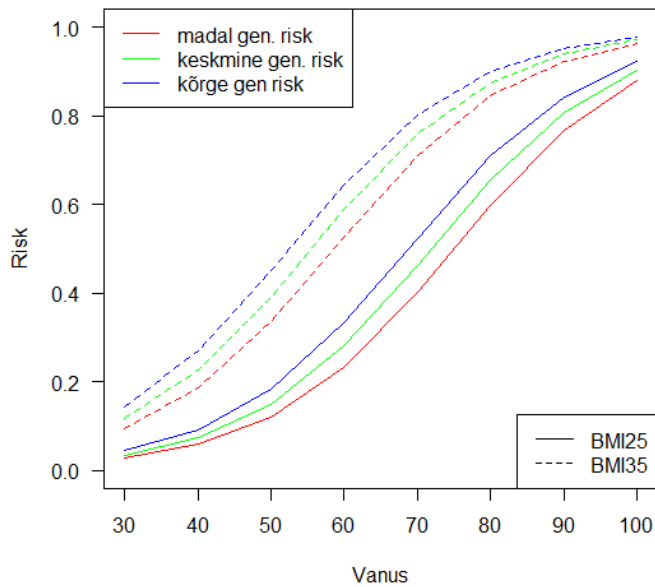
Järgnevatel joonistel vaadeldi erinevates geneetilistes riskigruppides, kuidas muutub risk haigestuda hüpertensiooni. Joonisel 2.7.1 on toodud eraldi meeste ja naiste risk haigestuda vanuse kasvades. Jooniselt on näha, et meestel on kõikidesse geneetilistesse riskigruppidesse kuulumisel suurem risk saada hüpertensiooni diagnoos, võrreldes naistega. Veel võib öelda, et vanuse kasvades suureneb risk haigestuda ning kõrgemasse geneetilisse riskigrupi kuulujatel on kõrgem risk saada hüpertensiooni diagnoos.



**Joonis 2.7.1** Riskigraafikud meestele ja naistele

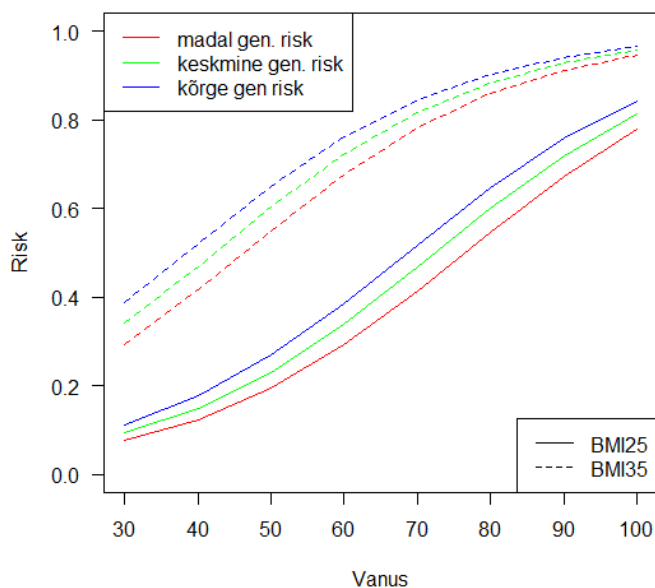
Riskigraafikud joonistati meestele ja naistele erinevate riski- ja kehamassiindeksi gruppide kaupa. Kehamassiindeks oli jaotud kaheks grupiks: kehamassiindeksiga 25, mis tähistab normaalkaalu kuuluvaid inimesi ning kehamassiindeksiga 35, mis tähistab tugevas ülekaalus olevaid inimesi.

Joonisel 2.7.2 on toodud riskigraafik naistele erinevates kehamassiindeksi- ja riskigruppidesse kuulumisel.



**Joonis 2.7.2.** Riskigraafik naistele kehamassiindeksi- ja riskigruppide kaupa

Jooniselt on näha, et rasvunud doonoritel on kõrgem risk saada hüpertooniatõve diagnoos võrreldes normaalkaalus olevate naissoost geenidoonoritega.. Ning samuti mida kõrgemas geneetilises riskigrupis geenidoonor on, seda suurem risk on saada hüpertooniatõve diagnoos. Joonisel 2.7.3 on toodud sarnane riskigraafik ja meeste jaoks.



**Joonis 2.7.3.** Riskigraafik meestele kehamassiindeksi- ja riskigruppide kaupa

Jooniselt on näha, et sarnaselt naistele on ka meestel nii, et rasvunudel on kõrgem risk haigestuda hüpertensiooni võrreldes normaalkaalus olevate meestega. Madalamasse geneetilisse riskigrupi kuulumise on ka risk madalam saada hüpertoonitõve diagnoos, keskmisesse geneetilisse riskigrupi kuulumisel on risk kõrgem ning neil geenidoonoritel, kes kuuluvad kõrgesse geneetilisse riskigrupi on risk saada hüpertensioon kõige kõrgem.

## KOKKUVÕTE

Käesoleva magistritöö eesmärgiks oli uurida ülegenoomsete assotsiatsiooniuringute põhjal leitud geenimarkerite prognoosivõimet hüpertooniatõve diagnoosi esinemise prognoosimisel ning sellest tulenevalt südame-veresoonkonnahaiguste riski hindamisel.

Kõrgvererõhutõbi on sageli esinev haigus, kuigi selle haiguse täpsed tekkepõhjused pole teada, on kindlasti oluline roll pärilikul eelsoodumusel ning mittetervislikel eluviisidel. Tuginedes Ajakirjas „*Nature*“ avaldatud artiklile (Ehret jt. 2011, 103), kus paljusid riike haarav uuring tegi kindlaks 29 geenimarkerit ehk SNP-d, mis osutusid seotuks vererõhu tasemega. Tehtud analüüsis ei osutunud suitsetamine oluliseks tunnuseks hüpertooniatõve prognoosimisel, mis võib tuleneda sellest, et diagnoosi saamisel on inimene otsustanud suitsetamisest loobuda. Olulisteks tunnusteks osutusid doonori vanus, sugu, kehamassiindeks ning geneetiline skoor, mis saadi, kui korrutati markeri väärtus ehk riskialleelide arv vastava beeta kordajaga artiklist „*Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk*“ (Ehret jt. 2011, 105) ning liideti saadud väärtused kokku üle 28 geenimarkeri, mis oli töö autoril kasutada. Geneetiline riskiskoor osutus väga oluliseks prognostiliseks tunnuseks logistilise regressioonimudeli põhjal ( $p < 0,001$ ). Samas, kasutades ROC-analüüsi, ilmnes et riskiskoori prognostiline väärtus on väike.

Esialgu tundus nagu geenianndmetel puuduski mõju hüpertooniatõve esinemise prognoosimisel. Geenianndmetega ja geenianndmeteta mudelitele joonistatud ROC-kõverad olid suures osas sarnased ning AUC erinevus väga väike. Kuna hüpertensiooni diagnoosi esinemine sõltub kahest otsustuspiirist: süstoolne vererõhk on üle 140 mmHg või diastoolne vererõhk on üle 90 mmHg, siis antud juhul seostusid geenimarkerid tugevamini vererõhu tasemega, kui diagnoosi esinemisega, mis võib olla see üheks põhjuseks, et ROC-kõverate erinevus on geenianndmeteta ja geenianndmetega mudelite puhul minimaalne.

On näha, et ümberklassifitseerimine töötab hüpertooniatõve prognoosimisel teatud gruppides, näiteks üle 60aastaste geenidoonorite seas, kus oli ka diagnoosiga doonoreid kõige rohkem: õiges suunas klassifitseerituid on rohkem, kui vales suunas klassifitseerituid ning vaadeldes protsente keskmises riskigrupis, mis on personaalse meditsiini seisukohalt kõige tähtsam grupp, on erinevus ligikaudu kahekordne.

Võrreldes geenidoonoreid sooliselt, siis meestel on risk saada hüpertooniatõve diagnoos kõrgem kui naistel ning kõrgemas geneetilise riskigrupi kuulujatel on kõrgem risk

saada hüpertensiooni diagnoos võrreldes nendega, kes kuuluvad keskmisesse või madalamasse geneetilisse riskigruppi.

Kui võrrelda geenidoonoreid kaalugruppide kaupa- normaalkaalus ja rasvunud geenidoonorid, siis kõrgem risk saada hüpertooniatõve diagnoos on neil doonoritel, kelle kehamassiindeks viitab tugevale ülekaalule. Meeste puhul oli näha, et madalamasse riskigruppi kuulujatel on madalam risk saada hüpertooniatõve diagnoos võrreldes keskmisesse ja kõrgemasse riskigruppi kuuluvate geenidoonoritega, sarnane tendents ilmnes ka naiste puhul.

Käesoleva töö puhul võib öelda, et pärilikul eelsoodumusel on mõju hüpertooniatõve diagnoosi esinemisele, kuid et täpsemat ülevaadet saada oleks mõistlik võtta vaatluse alla muude faktorite osas sarnased patsiendid, et näha geeniandmetest tulenevat lisainformatsiooni, kitsendusteta on suurem ja olulisem mõju indiviidi elustiilil. Praegusel hetkel on teada 29 geenimarkerit, mis on seotud vererõhu tasemega, antud töös kasutati neist 28, kuid teadmiste suurenemisel töötaksid käsitletud meetodid paremini, sest juba praegugi on mõne haiguse puhul teada üle 100 markeri ning geeniandmed annavad haiguse prognoosimisel uut informatsiooni.

# **Predictive ability of the genetic risk score for hypertension: analysis based on the data of the Estonian Genome Center, University of Tartu**

Master Thesis

Marja-Liisa Roos

## **Summary**

The aim of the master thesis was to study how well gene markers found by the genome-wide association study evaluate the probability of risk for hypertension.

In the analysis the gene donors' gender, age, and body mass index (BMI) were viewed, as well as the score of the diastolic and systolic blood pressure, as they were calculated in accordance to the context of hypertension.

In the analysis it seemed like genetic data is insignificant when evaluating hypertension. The ROC, drawn on the models with genealogical data and without the genetic data were mostly the same, and the difference of AUC was very little. As there are two cut-off points for the diagnosis of hypertension, systolic blood pressure is above 140 mmHg or diastolic blood pressure is above 90 mmHg, the gene markers' connection to the blood pressure level is stronger than the occurrence of the diagnosis. That, therefore, could be a reason why the difference between the ROCs on the models with and without the genetic data is minimal.

It can be seen that the genetic score does have an effect on the diagnosis of hypertension but it is rather weak compared to the other factors. When viewing the linear models that were created to assess the level of systolic and diastolic blood pressure it is clear that a higher genetic score causes the blood pressure to raise, and that especially among the gene donors who did not have hypertension. That is because it can be presumed that patients with the diagnosis are prescribed with medications that treat hypertension.

To avoid the influence of other factors a dataset was created that comprises gene donors who are 30-60 years old and whose BMI is between the range of 25-30. At first a control was found in accordance to every case and then models to evaluate hypertension were designed. Thus, having eliminated the influence of gender, age, and BMI it was found that there is a genetic influence of risk for hypertension.

When NRI between different groups were compared it was clear that gene data has a connection of the occurrence of hypertension. Its effect is best seen in the sub-dataset. It can also be seen that the effect of gene data when evaluating hypertension is bigger among women aged over 60 and among the men 60 and below. Therefore, it can be said that the gene markers' ability to evaluate risk for hypertension is distinguishable when the effect of gender, age, and BMI is eliminated.

When comparing the donors according to the gender then men have a higher risk of having hypertension than women and those who belong to a higher genetic risk group have a bigger risk of having it than those who belong to lower risk groups.

When comparing the gene donors according to their BMI it is the obese donors that have a higher risk to have hypertension than the donors in the normal BMI range.

## KASUTATUD KIRJANDUS

1. Ehret G.B., P.B. Munroe, P.B. Rice, M. Bochud jt. 2011. "Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk" *Nature* 478 (7367): 103-109. doi:10.1038/nature10405
2. Park, Seong Jo, Jim Mo Goo, Chan-Hee Jo. 2004. "Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists." *Korean Journal of Radiology* 5 (1): 11-18. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2698108/pdf/kjr-5-11.pdf>.
3. Greenhalgh Trisha. 1997. "How to read a paper: Papers that report diagnostic or screening tests." *BMJ* 1997 (8):315. doi: 10.1136/bmj.315.7107.540.
4. Cook, Nancy R.. 2008. "Statistical Evaluation of Prognostic versus Diagnostic Models: Beyond the ROC Curve." *Clinical Chemistry* 54 (1): 17-23. <http://www.clinchem.org/content/54/1/17.full.pdf+html>.
5. Campbell, Michael J., David Machin. 1999. *Medical Statistics: A Commonsense Approach*. New York: John Wiley & Sons.
6. Pencina, Michael J., Ralph B. D'Agostino Sr, Ralph B. D'Agostino Jr, Ramachandran S. Vasan. 2009. "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond." *Statistics in Medicine* 2008 (27): 157-172. doi:10.1002/sim.2929.
7. Pencina, Michael J., Ralph B. D'Agostino Sr, Ewout W. Steyerberg. 2011. "Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers." *Statistics in Medicine* 2011 (30): 11-21. doi: 10.1002/sim.4085.
8. Tartu Ülikooli Eesti Geenivaramu veebilehekülg. <http://www.geenivaramu.ee/index.php?id=70> (külastatud 24.02.2013)
9. Tervise Arengu Instituudi veebilehekülg. <http://www.toitumine.ee/korgvererohktobi-2/> (külastatud 24.02.2013)

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

Mina

Marja-Liisa Roos

---

*(autori nimi)*

(sünnikuupäev: 5.märts 1989)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose:  
Hüpertensiooni geneetilise riskiskoori prognoosivõime hindamine: analüüs Tartu Ülikooli Eesti Geenivaramu andmebaasi põhjal,
- 

*(lõputöö pealkiri)*

mille juhendaja on Krista Fischer, PhD

---

*(juhendaja nimi)*

- 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
- 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **20.05.2013**