

Tartu Ülikool

Loodus- ja täppisteaduste valdkond

Matemaatika ja statistika instituut

Enelin Haviko

Statiinide tarvitamise seos geneetiliste variantidega

TÜ Eesti Geenivaramu geenidoonoritel

Matemaatilise statistika eriala

Bakalaureusetöö (9 EAP)

Juhendajad:

Lili Milani, PhD

Krista Fischer, PhD

Tartu 2016

Statiinide tarvitamise seos geneetiliste variantidega TÜ Eesti Geenivaramu geenidoonoritel

Käesoleva bakalaureusetöö eesmärk on Tartu Ülikooli Eesti Geenivaramu geenidoonorite põhjal välja selgitada, kas invidiidid, kellel on kõrvaltoimetega seostatud genotüüp, jätvavad statiinide tarvitamise keskmisest sagedamini pooleli. Ühtlasi vaadeldakse, kas tausttunnused nagu vanus statiiniravi alustades ja kehamassiindeks mõjutavad ravimi tarvitamise katkestamist. Rakendatavad statistilise analüüsi meetodid on elukestusanalüüs, sealhulgas Coxi võrdeliste riskide mudel, ja logistiline regressioonanalüüs. Analüüsid, mis viidi läbi Tartu Ülikooli Eesti Geenivaramu ja Eesti Haigekassa digireseptide andmetega, tuvastasid, et SNP rs4363657 mõjutab statiiniravi katkestamise riski. Statiinide määramisel oleks seega soovitatav kasutada geneetilist testimist.

Märksõnad: farmakogeneetika, elukestusanalüüs, andmeanalüüs

CERCS teaduseriala: Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika (P160)

Association between genetic variants and use of statins among participants of the Estonian Genome Center, University of Tartu

The goal of this thesis is to use data from the Estonian Genome Center, University of Tartu, to find out whether individuals with genetic variants associated with side effects from statins treatment are more likely to discontinue their treatment. In addition, the thesis investigates whether background variables such as age at the beginning of statin treatment and body mass index affect discontinuation of treatment. For analysis of the data, methods of survival analysis, including Cox proportional hazards model, and logistic regression analysis are used. Analysis of the data from the Estonian Genome Center and digital prescriptions data from the Health Insurance Fund, detected a significant association between the genetic marker rs4363657 and the hazard for discontinuation of statin therapy. Therefore it is recommended to use genetic testing when prescribing statin therapy.

Keywords: pharmacogenetics, survival analysis, data analysis

CERCS research specialisation: Statistics, operation research, programming, actuarial mathematics (P160)

Sisukord

Sissejuhatus	6
1 Teoreetiline taust	7
1.1 Geneetiline varieeruvus	7
1.2 Statiinide farmakogeneetika	9
1.2.1 Olemus	9
1.2.2 Varasemad uuringud	10
1.3 Elukestusanalüüs	11
1.3.1 Põhimõisted	11
1.3.2 Kaplan-Meieri hinnang	14
1.3.3 Coxi võrdeliste riskide mudel	17
1.4 Logistiline regressioonanalüüs	18
2 Andmete kirjeldus	20
2.1 Digiretseptide andmed	20
2.2 Fenotüübiandmed	22
2.3 Genotüübiandmed	23
3 Analüüs	26
3.1 Analüüsimeetodite rakendamine tarkvaras R	26
3.2 Andmete jagamine gruppideks	26
3.3 Aeg ravi katkestamiseni	28
3.3.1 Elukestusanalüüs	30
3.3.2 Logistiline regressioonanalüüs	31
3.4 Aeg ravi katkestamiseni või statiini vahetuseni	32
3.4.1 Elukestusanalüüs	34
3.4.2 Logilistine regressioonanalüüs	35

Kokkuvõte	37
Viited.....	39
Lisad	42
Lisa 1. Coxi võrdeliste riskide mudeli hindamine tarkvaras R	42
Lisa 2. Elukestusfunktsiooni hindamine Kaplan-Meieri meetodiga tarkvaras R	43
Lisa 3. Logistilise regressioonmudeli hindamine tarkvaras R.....	45

Sissejuhatus

Personaalse genoomipõhise meditsiini väljatöötamiseks on vaja leida inimesele just tema genotüübiga sobivaid ravimeid. Teatud genotüübiga indiviididel võivad statiinid, mida kasutatakse kolesteroolitaseme alandamiseks, tekitada tõsiseid kõrvaltoimeid.

Käesoleva bakalaureusetöö eesmärk on 1119 Tartu Ülikooli Eesti Geenivaramu geenidoonori põhjal välja selgitada, kas kõrvaltoimetega seostatud genotüübiga indiviidid jätavad statiinide tarvitamise keskmisest sagedamini pooleli. Sealhulgas vaadeldakse, kas tausttunnused nagu vanus statiiniravi alustades ja kehamassiindeks mõjutavad ravimi tarvitamise katkestamist. Töö tulemusena oleks võimalik hinnata, kas Eestis peaks kasutama geneetilist testimist patsientidel, kellele soovitatakse määrata ravi statiinidega.

Esimeses peatükis antakse ülevaade geneetikast, varasematest uuringutest seoses statiinidega ning rakendatavatest statistilistest meetoditest. Teises osas kirjeldatakse kasutatavaid Tartu Ülikooli Eesti Geenivaramu ja Eesti Haigekassa digiretseptide andmeid. Kolmandas peatükis rakendatakse statistilisi meetodeid andmetele ja analüüsitakse tulemusi. Analüüsid on läbi viidud statistikatarkvara R [1].

Autor tänab juhendajaid Lili Milanit ja Krista Fischerit huvitava temapüstituse ja rohkete nõuannete eest.

1 Teoreetiline taust

Esimese peatüki eesmärk on lugejale tutvustada nii töö geneetilist kui ka statistilist tausta. Esimesed kaks alapeatükki kirjeldavad, kuidas on inimesed üksteisest erinevad ning miks on seda ravimite, sh statiinide, määramisel kasulik teada. Viimased alapeatükid annavad ülevaate kahest statistilise analüüsi meetodist, mis sobivad iseloomustama statiinide tarvitamise seost geneetiliste variantidega.

1.1 Geneetiline varieeruvus

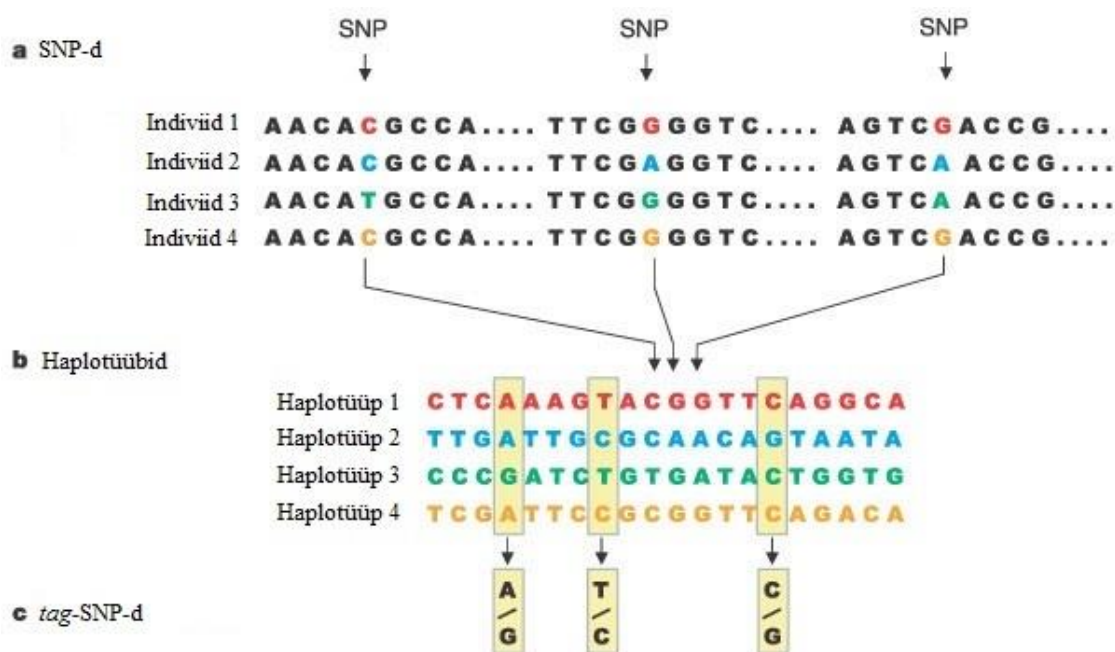
Järgnev alapeatükk põhineb A. Heinaru raamatul *Geneetika* [2], v.a seal, kus on märgitud teisiti.

Elusorganismi ehituskivideks on rakud, mis koosnevad mitmetest eritüübilistest molekulidest nagu lipiidid, valgud jms. Sealhulgas sisaldab iga rakk geneetilist materjali, mida tuleb võimalikult muutumatult säilitada ja kanda edasi järglasrakkudesse. Geneetiline info on rajatud kromosoomidesse, mis igas rakus moodustavad kaks 23 kromosoomist koosnevat geenikomplekti, kus üks on saadud emalt ja teine isalt. Geen on lõik DNA-st, mis kodeerib valku. Kokku pakitud DNA moodustab kromosoomi, mis koosneb kahest komplementaarsest nukleotiidide ahelast. Nukleotiide on nelja erinevat tüüpi: adeniin (A), tümiin (T), tsütosiin (C) ja guaniin (G). Komplementaarsus tähendab olukorda, kus kaks ahelat täiendavad end vastastikku – DNA molekulis ühe ahela tümiin (T) paardub teise ahela adeniiniga (A) ja ühe ahela tsütosiin (C) teise ahela guaniiniga (G). Sekvencerides DNA ühe ahela saab komplementaarsuse alusel teada ka teise ahela nukleotiidse järjestuse. Hinnanguliselt on inimese genoomi kogupikkus umbes kolm miljardit nukleotiidipaari.

Alleel on kromosoomi määratud piirkonnas ehk lookuses olev üks kahest või mitmest alternatiivsest geenivariandist. Alleelide koosseis moodustab indiviidi genotüübi vastavas lookuses. Isik on homosügootse genotüübiga, kui geeni mõlemas lookuses on sama alleel, ja heterosügootse genotüübiga, kui alleelipaaris on erinevad alleelid. Homo- ja heterosügootse genotüübiga indiviide nimetatakse vastavalt homosügootiks ja heterosügootiks.

Ühenukleotiidne polümorfism ehk SNP (*single nucleotide polymorphism, SNP*) on DNA järjestuse varieeruvus, mis väljendub ühe nukleotiidi muutumisel genoomis. SNP-d sageli ei põhjusta kodeeritavate aminohapete (valgu koostisosa) muutust, mistõttu ei ole nad alati organismile kahjulikud.

SNP-sid esineb keskmiselt üks 200-300 aluspaari kohta ning nende kombinatsiooni ühes kromosoomiosas nimetatakse haplotüübiks. Konkreetse indiviidi haplotüübiploki järjestuse kindlakstegemiseks piisab vaid mõne iseloomuliku SNP väljaselgitamisest. Selliseid SNP-sid nimetatakse *tag*-SNP-deks. *Tag*-SNP-d esindavad teisi samas piirkonnas asuvaid SNP-sid, kuna on nendega aheldunud. Alleelid on aheldunud, kui nad päranduvad suure tõenäosusega koos, mille tagab nende lähedikkude paiknemine. Joonis 1.1 kirjeldab SNP-de olemust täpsemalt.



Joonis 1.1 SNP-d, haplotüübid ja *tag*-SNP-d [3]. a) Näha nelja inimese kromosoomi samad ahelduspiirkonnad. Enamus DNA järjestusest on neil identne, välja on toodud kolm varieeruvat nukleotiidi. Igal SNP-l on kaks võimalikku alleeli – esimesel C ja T, teisel ning kolmandal G ja A. b) Isikute haplotüübid määratud ahelduspiirkonnas. c) On toodud kolm *tag*-SNP-d, mille abil on võimalik määrata geneetilist varieeruvust, ilma et oleks vaja genotüüpiseerida haplotüübi kõiki SNP-sid. Näiteks, kui konkreetsel kromosoomil oli märgistatud SNP-de suhtes järjestus A-T-C, siis see läheb kokku esimese haplotüübiga.

Ühenukleotiidsed polümorfismid on kõige sagedasem inimestevaheline geneetiline varieeruvus. Geneetilise polümorfismi tõttu ongi inimesed unikaalsed – sarnasus kahe inimese DNA vahel on 99.9%, kuid SNP-de poolt põhjustatud 0.1% erinevus tagab geneetilise ainulaadsuse.

Lisaks eelnevale võivad variatsioonid DNA-s määrata, kuidas inimestel arenevad haigused, kuidas reageeritakse kemikaalidele, ravimitele ja vaktsiinidele ning millised on nende kõrvaltoimed. Seega on SNP-d kasulikud haiguste diagnoosimisel, ravimite määramisel ning arendamisel. Meditsiinalane uurimistegevus keskendubki järjest enam SNP-dele, püüdes välja arendada personaalset meditsiini, mis näib olevat üks tulevikusuundadest. [4]

1.2 Statiinide farmakogeneetika

1.2.1 Olemus

Farmakogeneetika on võrdlemisi uus teadusharu, mis uurib, kuidas geenid mõjutavad inimese ravimivastust, arendades sellega efektiivset ja ohutut ravimi annustamist [5]. Enne statiinide ja nende farmakogeneetika kirjeldamist, tutvustatakse järgnevas lõigus põgusalt kolesterooli olemust.

Kolesterool kuulub lipiidide rühma ja on inimorganismi elutegevuse oluline koostisosa, reguleerides rakkude talitust. Kolesterool on paljude hormoonide koostises, see võtab osa rasvade ainevahetusest ning sapphapete ja D-vitamiini sünteesist. Organismis oleva kolesterooli saab jagada kaheks: ligikaudu 70% sünteesitakse kehaomaselt (peamiselt maksas) ja umbes 30% saadakse söödavast toidust. Eristatakse „head“ ehk HDL-kolesterooli ja „halba“ ehk LDL-kolesterooli. HDL-kolesterool aitab vabaneda liigsest kolesteroolist, transportides selle maksa, kus toimub ümbertöötlemine või organismist väljutamine. LDL-kolesteroolil on omadus kleepuda veresoonte seintele, mis aja möödudes põhjustab veresoonte lupjumist ehk ateroskleroosi. Üldkolesterooli norm on väiksem kui 5.0 mmol/L, sh HDL-kolesterool suurem kui 1.1 mmol/L ja LDL-kolesterool väiksem kui 3 mmol/L. Kõrge kolesteroolisisaldus veres on üheks kardiovaskulaarsete ehk südame-veresoonkonna haiguste põhjustajaks. [6]

Statiinid on kõige laialdasemalt välja kirjutatud südame-veresoonkonna haiguste ravimid, mis toimivad peamiselt vereplasma LDL-kolesterooli alandades. Seda, kuidas geenid mõjutavad inimestevahelist erinevust statiinide toimes, nimetatakse statiinide farmakogeneetikaks. Statiinide farmakogeneetika keskendub põhiliselt kahele aspektile:

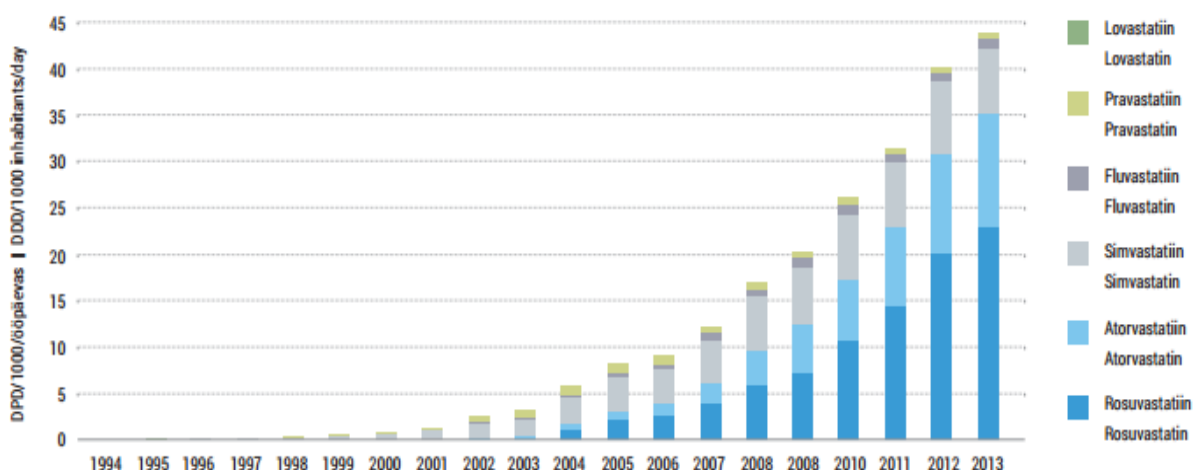
- 1) efektiivsus – millisel määral LDL-kolesterooli alandatakse ning südamelihase infarkt ja muud olulised tüsistused ära hoitakse;
- 2) müotoksilisus (*myotoxicity*) – kui mürgine on ravim lihastele. [7]

Müotoksilised ravimid võivad põhjustada müopaatiat [8]. Müopaatia on üldnimetus lihase funktsiooni häirele, mille korral lihaskiudude normaalne talitus on erinevatel põhjustel häiritud. Sellega kaasnevad tihti lihassenõrkus, -krambid ja -jäikus [9].

1.2.2 Varasemad uuringud

Massachusetts Meditsiini Seltsi 2008. aasta uuringus selgus, et geen *SLCO1B1* on kaks SNP-d (rs4363657 ja rs4149056), mille puhul teatud genotüübiga indiviididel on statiiniravis suurem šans müopaatiale, eriti suures annuses simvastatiini puhul. Šansside suhe müopaatiale oli ühe C alleeliga inimeste puhul 4.5 ja kahe C alleeliga inimeste puhul 16.9 võrreldes homosügootiga TT [10]. Aasta hiljem tõestasid soomlased, et geeni *ABCG2* SNP rs2231142 mõjutab fluva- ja simvastatiini imendumist. Genotüübiga AA patsientidel jaotus ravim kehas keskmiselt suuremale alale kui CA või CC genotüübiga indiviididel [11]. 2012. aastal Koreas läbiviidud uuring näitas, et isikutel, kellel oli geen *HMGCR* SNP-s rs3846662 genotüüp GG, oli atorvastatiini võtmise järgselt LDL-kolesterooli keskmiselt 18-33% rohkem kui isikutel genotüübiga AA [12]. Samal ajal leiti Ameerikas, et pravastatiini ravi korral vähendab *SLCO1B1* geenis olev SNP rs4149056 vanematel inimestel (keskmine iga 75.3) LDL-kolesterooli erinevalt. Metsiktüüpi alleelipaari TT puhul vähenes LDL-kolesterool 37%, CT korral 36% ning CC korral 31.8% [13]. 2012. aasta lõpuks näidati 247 Hiina patsiendi puhul, et samas geen *SLCO1B1* olev SNP rs4149081 mõjutab rosuva- ja simvastatiiniravi korral kolesteroolitaseme alanemist. Rs4149081-s genotüübiga AA ja AG inimestel vähenes LDL-kolesterool rosuva- ja simvastatiiniga vastavalt 4.6 ja 4.0% rohkem kui metsiktüüpi alleelidega (GG) inimestel [14].

Statiinide kasutamine on Eestis aastatega järjest suurenenud (vt joonis 1.2). 2013. aastaks oli tarvitamise intensiivsus kasvanud 43.65 defineeritud päevadoosini tuhande inimese kohta ööpäevas. Eestis kasutatakse peamiselt kuut erinevat statiini ning nende hulgast kirjutatakse kõige rohkem välja rosuvastatiini. Statiinide tarvitajate arv kasvab, mistõttu on vajadus nende ravimite mõju hindamiseks põhjendatud.



Joonis 1.2 Statiinide kasutamine Eestis 1994-2013 [15, lk 55]

1.3 Elukestusanalüüs

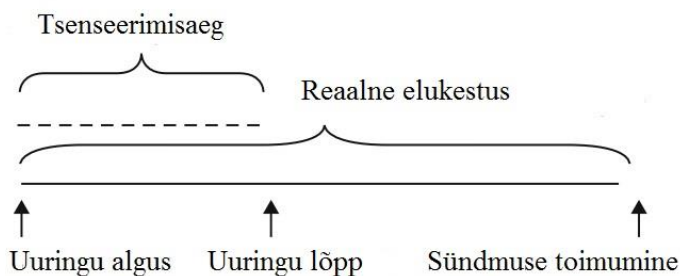
1.3.1 Põhimõisted

Elukestusanalüüs (*survival analysis*) on kogum statistilistest protseduuridest, kus huvipakkuvaks tunnuseks on aeg sündmuse toimumiseni. Aja all mõistetakse aastaid, kuid, nädalaid või päevi kindlaks määratud algmomendist kuni teatud sündmuse esinemiseni. Sõltuvalt valdkonnast mõistetakse sündmuse all näiteks surma, tervenemist, seadme riknemist, abielulahutust, ravi katkestamist jms. [16]

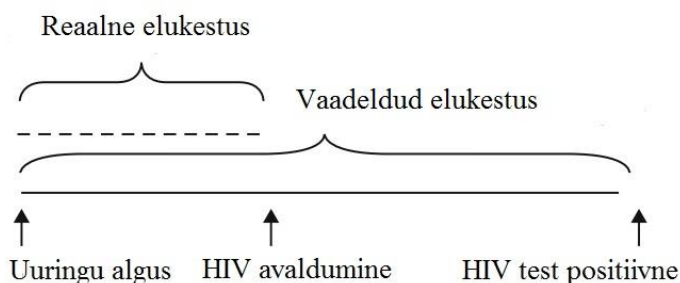
Aega fikseeritud algmomendist kuni sündmuse toimumiseni nimetatakse elukestuseks (*survival time*), vastavat juhuslikku suurust tähistatakse suure T -ga ning realiseerinud väärtus väikese t -ga. Kui huvipakkuv sündmus pole subjekti jaoks uuringu jooksul toimunud, siis öeldakse, et subjekti elukestus on tsenseeritud ning subjekti jälgimisaega C nimetatakse tsenseerimisajaks. Indikaator-tunnus $\delta = I(T < C)$ omab väärtust 1, kui uuringu vältel subjektile esines huvipakkuv sündmus ja 0, kui subjekti elukestus on

tsenseeritud ehk sündmus ei ole vaadeldud. Viimasel võib olla kolm põhjust: sündmus pole uuringu vältel toimunud, subjekt lahkus uuringust enne katse lõppu (seniks polnud sündmus toimunud) või suri katse jooksul, kuid põhjus ei ole seotud uuringuga. [16,17] Näiteks, olgu läbi viidud mingi meditsiiniline uuring, mis kestis üks aasta. Patsient A selle aja jooksul ei haigestunud, seega tema elukestus on tsenseeritud ja tsenseerimisaeg on üks aasta, reaalne elukestus on uurijatele teadmata. Patsient B liitus uuringuga poole pealt ning haigestus vaatluse lõpus. Tema elukestus on kuus kuud, tsenseerimist ei esine.

Kestusandmeid on võimalik tsenseerida paremalt ja vasakult. Kõige rohkem esineb eelpool kirjeldatud parem-tsenseerimist. See tähendab, et subjekti täpsel elukestusel niioelda lõigatakse ära parem pool, kui uuring lõpeb, subjekt loobub uuringus osalemast või sureb. Seega on tsenseeritud subjekti tsenseerimisaeg lühem kui uurijatele teadmata reaalne elukestus (vt joonis 1.3). Vasakult tsenseeritud andmete puhul on subjekti vaadeldud elukestus pikem kui reaalne elukestus, seega huvipakkuv sündmus toimus enne selle fikseerimist. [16] Näide on toodud joonisel 1.4, kus huvipakkuvaks sündmuseks on positiivne HIV testi tulemus.



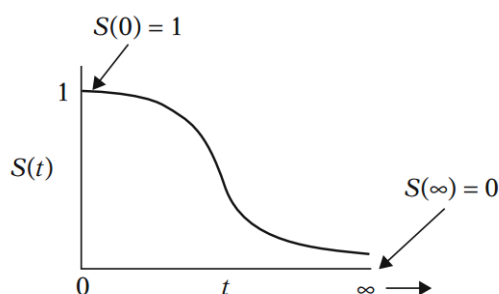
Joonis 1.3 Paremt tsenseerimine



Joonis 1.4 Vasakult tsenseerimine [16, lk 8]

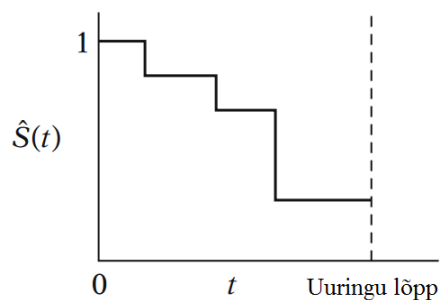
Elukestus- ehk teisisõnu üleelamisfunktsioon (*survival function*) on tõenäosusfunktsioon, mis näitab millise tõenäosusega subjekti elukestus on suurem kui t : $S(t) = P(T > t)$ [17]. Teoreetiliselt on t muutumispiirkond $(0, \infty)$ ja elukestusfunktsiooni saab joonistada kui sileda kõvera (vt joonis 1.5). Lisaks on funktsioonil järgnevad omadused [16].

1. $S(t)$ on mittekasvav funktsioon.
2. $S(0) = 1$ ehk uuringu alguses ($t = 0$) on subjekti elukestuse, mis on suurem kui 0, tõenäosus 1.
3. $S(\infty) = 0$ ehk uuringu perioodi pikaajalisel kasvamisel ($t = \infty$) ei ole enam ühtegi subjekti elus ning elukestusfunktsioon saab väärtuse 0.



Joonis 1.5 Teoreetiline elukestusfunktsioon [16, lk 9]

Reaalse andmestiku puhul saadakse elukestusfunktsioonile hinnang $\hat{S}(t)$, mis on enamasti treppfunktsiooni kujul. Kuna uuringu ajaline pikkus on piiratud, tähendab see, et igal subjektil ei esine tingimata antud ajavahemikus huvipakkuvat sündmust, mistõttu ei pruugi ka $\hat{S}(t)$ -l olla nullilähedaseid väärtusi [16]. Näide elukestusfunktsiooni hinnangust on joonisel 1.6.



Joonis 1.6 Elukestusfunktsiooni hinnang [16, lk 9]

Lisaks elukestusfunktsioonile on kasutusel ka riskifunktsioon (*hazard function*) h , mis näitab sündmuse toimumise tõenäosust ajamomendil t , tingimusel et sündmus enne ajamomenti t ei toimunud [17]:

- 1) $h(t_j) = h_j = P(T = t_j | T \geq t_j)$, kus $t_j, j \in (1, 2 \dots)$, on diskreetsed ajahetked, näiteks päevad;
- 2) $h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$, kus aeg on esitatud pideva tunnusena.

Diskreetse ajatunnuse korral kehtib $S(t_j)$ ja h_j vahel seos [17]:

$$S(t_j) = \prod_{j: 0 < t_j < t} (1 - h_j).$$

1.3.2 Kaplan-Meieri hinnang

Järgnev alapeatükk põhineb K. Fischeri loengukonspektil *Elukestusanalüüs* [17].

Tsenseeritud andmete korral kasutatakse enim elukestusfunktsiooni hindamisel Kaplan-Meieri hinnangut. Selleks kirjutatakse andmestik välja nagu tabelis 1.7.

Tabel 1.7 Näide andmestikust: t_j – sündmuse või tsenseerimise toimumise aeg, d_j – sündmuste arv vastaval ajamomendil, c_j – tsenseeritute arv sellel ajamomendil, r_j riskigrupi suurus vahetult enne ajamomenti t_j (riskigrupi kuuluvad need, kellel vahetult enne momenti t_j sündmus pole toimunud või kes pole vahetult enne momenti t_j tsenseeritud).

t_j	d_j	c_j	r_j
1	5	0	20
2	3	0	15
5	2	0	12
7	0	1	10
10	1	0	9
12	1	1	8
16	1	0	6
20	0	1	5
24	1	0	4
30	0	2	3
32	0	1	1

Tabelis 1.7 on t_j mõõdetud diskreetse tunnuseks, näiteks päevades katsesse sisenemise hetkest. Meeles tuleb pidada, et t_j tähistab siin jälgimisaja pikkust. Seega, kui ajamomendil t_j toimus kaks või enam sündmust, siis see tähendab, et need subjektid olid vaatluse all sama kaua, kuid sündmused ei pruukinud tegelikult toimuda samal kuupäeval. Tabelist on näha, et 14 subjektile toimus huvipakkuv sündmus ja 6 subjekti olid tsenseeritud. Seega vaatluse all oli 20 subjekti.

Tabeli põhjal saab hinnata riski ehk riskifunktsiooni väärtust ajamomendil t_j :

$$\hat{h}_j = \frac{d_j}{r_j}.$$

Vaadates huvipakkuva sündmusena surma, näitab see arv, kui suur osa ajal t_j elusolijatest sureb sel ajahetkel.

Kasutades seost riskifunktsiooni ja elukestusfunktsiooni vahel, saab leida Kaplan-Meieri hinnangu elukestusfunktsioonile:

$$\hat{S}(t_j) = \prod_{j:0 < t_j < t} (1 - \hat{h}_j).$$

Selle hinnangu puhul piisab, kui arvutada riskifunktsiooni väärtus ainult nendel ajahetkedel, kui toimus sündmus või tsenseerimine. Näiteandmestikus toimus sündmus üks päev pärast katsega liitumist viiel subjektile. Ajahetkel 1 on riskifunktsiooni väärtus $\frac{5}{20} = 0.25$ ja elukestusfunktsiooni väärtus $1 - 0.25 = 0.75$. Kaks päeva pärast katsesse sisenemist toimus sündmus kolmel subjektile. Riskifunktsiooni väärtus on sellel ajahetkel $\frac{3}{15} = 0.20$ ja elukestusfunktsiooni väärtus $0.75(1 - 0.20) = 0.6$. Näiteandmestik koos kõigi 11 hinnanguga riski- ja elukestusfunktsioonile on toodud tabelis 1.8.

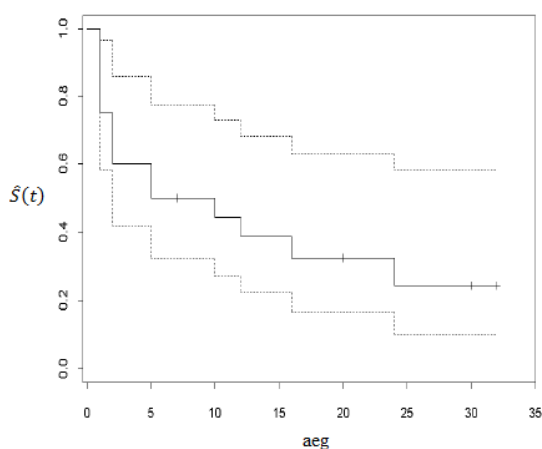
Tabel 1.8 Näiteandmestik koos hinnangutega riski- ja elukestusfunktsioonile

t_j	d_j	c_j	r_j	\hat{h}_j	$\hat{S}(t_j)$
1	5	0	20	0.25	0.75
2	3	0	15	0.20	0.60
5	2	0	12	0.17	0.50
7	0	1	10	0	0.50
10	1	0	9	0.11	0.44
12	1	1	8	0.13	0.39
16	1	0	6	0.17	0.32
20	0	1	5	0	0.32
24	1	0	4	0.25	0.24
30	0	2	3	0	0.24
32	0	1	1	0	0.24

Tabeli 1.8 põhjal saab hinnata ka elukestuse mediaani ehk ajahetke, milleks pooltel vaatluse all olevatest subjektidest sündmus on toimunud:

$$\widehat{Med}(T_i) = \min\{t_i | \hat{S}(t_i) \leq 0.5\}.$$

Vahel võetakse elukestuse mediaani hinnanguks ka sellise lõigu keskmine, kus $\hat{S}(t) = 0.5$. Antud näite puhul on selliseks hinnanguks 7.5 kuud (5-st kuni 10. kuuni $\hat{S}(t) = 0.5$, seega keskmine on $\frac{5+10}{2} = 7.5$). Kaplan-Meieri hinnangule saab leida ka standardvea ja selle põhjal usaldusintervalli. Näiteandmestiku põhjal saadud elukestusfunktsiooni graafik koos 95% usaldusintervalliga on kujutatud joonisel 1.9.



Joonis 1.9 Kaplan-Meieri hinnang elukestusfunktsioonile koos 95% usalduspiiridega (katkendlik joon)

1.3.3 Coxi võrdeliste riskide mudel

Järgnev alapeatükk põhineb K. Fischeri loengukonspektil *Elukestusanalüüs* [17], v.a seal, kus on märgitud teisiti.

Coxi võrdeliste riskide mudel (*Cox proportional hazards model*) on üks populaarsemaid mudeleid elukestussanalüüsis. Selle eesmärk on leida seos elukestuse ja seletavate tunnuste vahel ning nende põhjal hinnata huvipakkuva sündmuse toimumise riski [18]. Mudeli kuju on järgmine:

$$h(t|\mathbf{X}) = h_0(t)\psi(\mathbf{X}) = h_0(t)e^{\mathbf{X}\boldsymbol{\beta}},$$

kus t on ajahetk, millele riski arvutatakse, $\mathbf{X} = (X_1, X_2, \dots, X_p)$ on argumenttunnuste vektor, $h_0(t)$ on baasriskifunktsioon ja $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ on regressioonikordajate ehk parameetrite veeruvektor.

Tuleb panna tähele, et baasriskifunktsioon sõltub ajast t , kuid mitte argumenttunnustest X_m , $m = 1, 2, \dots, p$. Seevastu eksponentosa sõltub argumenttunnustest, mitte aga ajast t ehk argumenttunnuste mõjud on ajast sõltumatud. Huvi pakubki eksponentosas parameetrite hindamine, milleks kasutatakse osalise tõepära meetodit, mida käsitletakse analoogselt suurima tõepära meetodiga. Osaliseks nimetatakse seda sellepärast, et Coxi mudeli puhul informatsioon $\boldsymbol{\beta}$ kohta sisaldub sündmuste toimumiste järjekorras, mitte aga konkreetses ajamomentides. Tähistagugi $\tau_1 < \tau_2 < \dots < \tau_d$ erinevaid ajamomente, mil toimus huvipakkuv sündmus. Olgu i_j selline subjekt, kellel esines sündmus momendil τ_j ning olgu t_i , $i = 1, 2, \dots, n$, i -nda subjekti elukestus. Seega riskigrupp ajamomendil τ_j on $R(\tau_j) = \{i: t_i \geq \tau_j\}$. Tähistagu \mathbf{x}'_i i -nda subjekti argumentide vektori \mathbf{X}_i väärtuste vektorit. Kui i -nda subjekti risk momendil τ_j on $h(\tau_j, \mathbf{x}'_i)$ ja $\psi(i) = \psi(\mathbf{x}'_i, \boldsymbol{\beta})$, siis osalise tõepära funktsioon avaldub kujul:

$$L = \prod_{j=1}^d \frac{\psi(i)}{\sum_{k \in R(\tau_j)} \psi(k)}.$$

Logaritmiline tõepärafunktsioon $l = \ln(L)$ avaldub $\psi(i) = e^{x' i \beta}$ korral kujul:

$$l = \sum_{j=1}^d x'_{ij} \beta - \sum_{j=1}^d \ln \sum_{k \in R(\tau_j)} e^{x'_{kj} \beta} .$$

Võttes sellest funktsioonist esimesed tuletised β_m -de, $m = 1, 2, \dots, p$, järgi ning võrdsustades need nulliga, saab leida parameetrite hinnangud, kasutades näiteks Newton-Raphsoni iteratsioonimeetodit. Positiivne regressioonikordaja suurendab ja negatiivne vähendab argumenttunnuse suurenedes sündmuse toimumise riski [18].

Eelnev arutelu osalise tõepära funktsiooni leidmiseks kehtib, kui igas ajaühikus toimub vaid üks sündmus. Kui sündmusi on mingitel ajamomendil rohkem kui üks, siis tuleb osalise tõepära funktsiooni modifitseerida. Statistikapakett R pakub kolme meetodit: Efron (vaikimisi), Breslow, „täpne osaline tõepära“ [19].

Parameetrite olulisust hinnatakse enamasti Waldi statistiku abil, mis avaldub kujul $Z = \frac{\hat{\beta}}{SE(\hat{\beta})}$ ning nullhüpoteesi ($\beta = 0$) kehtides normaaljaotusega [16]. R väljastab Coxi mudeli korral automaatselt Waldi statistiku ja sellele vastava p-väärtuse.

Coxi võrdeliste riskide mudeli eelduseks on, et argumenti erinevatele väärtustele vastavad riskid on võrdelised. Seda eeldust saab programmis R kontrollida funktsiooniga `cox.zph()`, kus argumentiks on hinnatud Coxi VR mudel, mille eelduse täidetust soovitakse kontrollida [20].

1.4 Logistiline regressioonanalüüs

Järgnev alapeatükk põhineb E. Kääriku loengukonspektil *Andmeanalüüs II* [21].

Logistilise regressioonimudeli korral on uuritava tunnuse võimalikud väärtused 1 ja 0, kus 1 tähistab huvipakkuva sündmuse Y esinemist. Huvi pakub seos uuritava tunnuse esinemise tõenäosuse $\pi = P(Y = 1)$ ja argumenttunnuste vahel. Kuna prognoositakse tõenäosust, mis on tõkestatud lõigul $[0,1]$, on mõistlikum leida teisendus kogu reaalteljele. Selleks sobib *Logit* seosefunktsioon $logit(\pi) = \ln \frac{\pi}{1-\pi}$. Logistilise mudeliga hinnatakse seega šansi logaritmi

$$\ln \frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p,$$

π on sündmuse esinemise tõenäosus, $\frac{\pi}{1-\pi}$ on sündmuse esinemise šanss, $\beta_0, \beta_1, \dots, \beta_p$ on tundmatud parameetrid ja x_1, x_2, \dots, x_p on argumenttunnused.

Mudeli parameetrite β_m , $m = 0, 1, \dots, p$, hindamine toimub suurima tõepära meetodil ning nende olulisust hinnatakse nagu Coxi võrdeliste riskide mudeli korral enamasti Waldi statistikuga. Parameetrite hinnanguid interpreteeritakse šansside suhte muutuse abil. Kahe isiku, i -nda ja j -nda, šansside suhe on

$$OR = \frac{\frac{\pi_i}{1-\pi_i}}{\frac{\pi_j}{1-\pi_j}}.$$

Kui argument x_m , $m = 1, 2, \dots, p$, muutub c ühiku võrra, siis šansside suhe muutub $e^{c\hat{\beta}_m}$ korda, tingimusel et teised argumentid jäävad samaks. Kui parameetri β usaldusvahemik on (a, b) , siis vastava šansside suhte usaldusvahemik on (e^a, e^b) . Šansside suhe pole oluline, kui 1 sisaldub usaldusvahemikus (e^a, e^b) . Kui aga mõlemad usaldusvahemiku otspunktid on ühest suuremad, siis on tegu riskiteguriga.

2 Andmete kirjeldus

Teise peatüki eesmärk on anda ülevaade kasutatud andmetest. Vaatluse all oli 1119 TÜ Eesti Geenivaramu geenidoonorit, kes tarvitasid perioodil 4. jaanuar 2010 kuni 4. august 2015 statiine. Nende kohta olid saadaval Eesti Haigekassa statiinide digiretseptide andmed ning TÜ Eesti Geenivaramu poolt kogutud fenotüübi- ja genotüübiandmed. Geenivaramuga liitusid vaatlusalused perioodil 15. oktoober 2002 kuni 6. juuni 2013. Seega vaid osa inimestest liitus geenivaramuga uuritava perioodi 4. jaanuar 2010 kuni 4. august 2015 jooksul. Digiretseptide andmed olid hilisematel liitunudtel ikkagi saadaval alates 4. jaanuarist 2010.

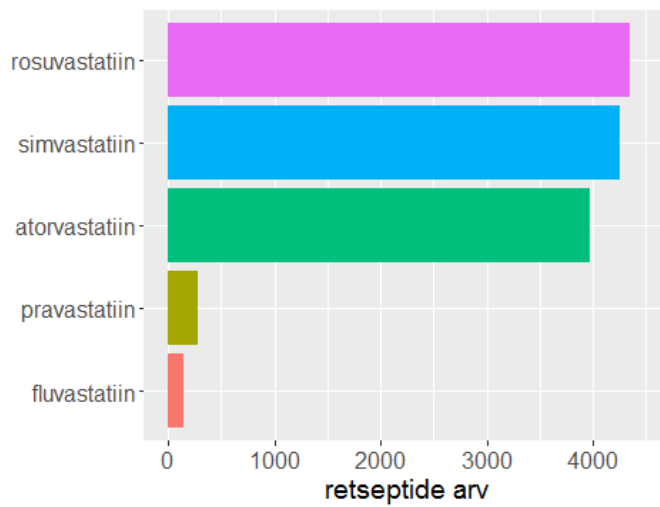
2.1 Digiretseptide andmed

Vaadeldud viie aasta ja seitsme kuu jooksul realiseeriti vaatlusaluste seas kokku 12 999 statiiniretsepti. Digiretseptide andmestikust pakkus huvi kaks tunnust: ravimi ATC-kood (*real_atc_code*) ja retsepti realiseerimiskuupäev (*real_kpv*). Lisaks kasutati indiviidi identifitseerivat koodi, et eristada samale isikule kirjutatud retsepte ning ühendada retseptide andmestik geenivaramu fenotüübi- ja genotüübiandmestikega.

Selleks, et ravimite tarvitamist oleks parem analüüsida ja seeläbi tarvitamise kvaliteeti tõsta, klassifitseeritakse neid ATC-koodide (*Anatomical Therapeutic Chemical codes*) abil. Grupeerimisel arvestatakse, millist elundkonda ravim mõjutab ning millised on ravimi keemilised, farmakoloogilised ja terapeutilised omadused [22]. Töös kasutatud andmestikus esines viis erinevat statiini, nende nimed koos ATC-koodidega on koondatud tabelisse 2.1. Jooniselt 2.2 on näha, et 12 999 retsepti seas oli kõige rohkem rosuvastatiini ja kõige vähem fluvastatiini retsepte.

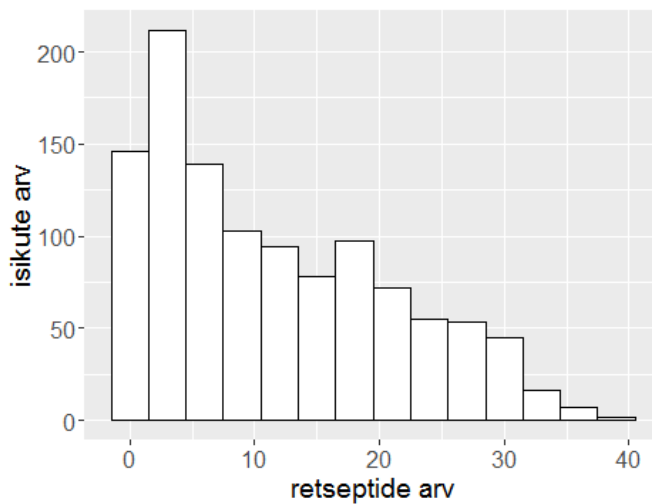
Tabel 2.1 ATC-koodid ja nende tähendus

ATC-kood	C10AA01	C10AA03	C10AA04	C10AA05	C10AA07
ravim	simvastatiin	pravastatiin	fluvastatiin	atorvastatiin	rosuvastatiin



Joonis 2.2 Realiseeritud retseptide jaotus vastavalt statiini tüübile

Uuritud geenidoonorid realiseerisid ühe või enam retsepti. Keskmiselt võttis üks inimene välja $11.6 \approx 12$ ($SD = 9.4$) retsepti. Täpsemalt saab realiseeritud retseptide arvu jaotust isiku kohta vaadata jooniselt 2.3.



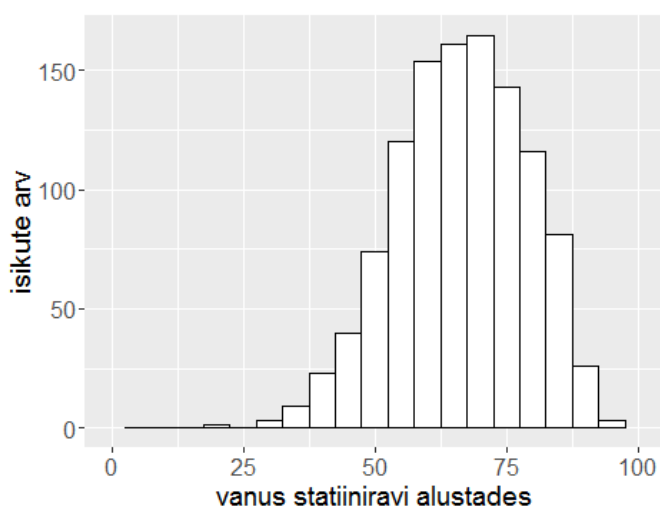
Joonis 2.3 Realiseeritud retseptide arvu jaotus isiku kohta

2.2 Fenotüübiandmed

Geenidonorite fenotüübiandmetest pakkus huvi viis tunnust:

- geenivaramuga liitumise kuupäev (*liitkuup*);
- vanus geenivaramuga liitudes (*vanus*);
- isiku surma kuupäev või viimane kuupäev, mil oli teada, et isik oli elus (*vkuup*);
- surma staatus, kus 0 – elus, 1 – surnud (*surnud*);
- kehamassiindeks geenivaramuga liitudes, kg/m² (*bmi*);

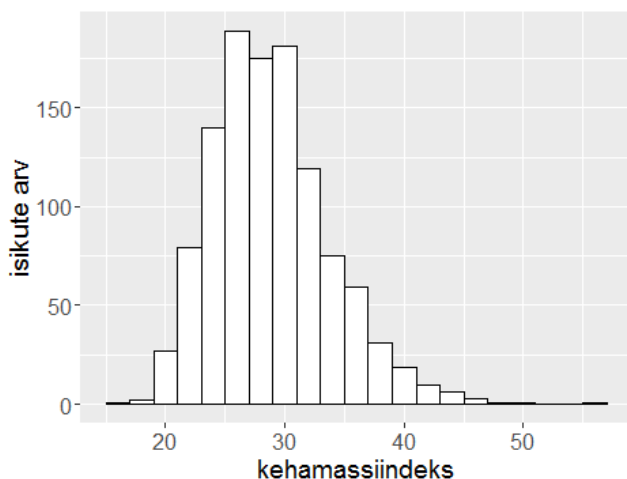
Kasutades digiretseptide andmetest esimese statiiniresepti realiseerimise kuupäeva ja fenotüübiandmetest tunnuseid *vanus* ja *liitkuup*, sai arvutada isiku vanuse statiiniravi alustades (uus tunnus *vanus_statiinid*). Kõige noorem raviga alustanu oli 22- ning vanim 96-aastane. Keskmiselt alustati statiiniraviga vanuses $66.3 \approx 66$ ($SD = 12.2$) aastat. Indiviidide vanusjaotus statiiniravi alustades on joonisel 2.4.



Joonis 2.4 Indiviidide vanusjaotus statiiniravi alustades

Geenivaramu andmebaasi viimane linkimine Eesti surmapõhjuste registri andmebaasiga toimus 10. septembril 2015. Selle põhjal defineeriti tunnus *surnud* väärtusega 1, kui isik oli selleks ajaks surnud. Kui *surnud* väärtus oli 0, siis isik enne 10. septembrit 2015 ei surnud. Alates 4. jaanuar 2010 kuni 10. september 2015 suri vaadeldud andmestikus 130 geenidonorit.

Vaatlusaluste keskmine kehamassiindeks geenivaramuga liitudes oli $28.9 \approx 29$ ($SD = 5.0$) kg/m^2 . Täpsemalt saab kehamassiindeksi jaotust vaadata jooniselt 2.5.



Joonis 2.5 Indiviidide kehamassiindeksi jaotus geenivaramuga liitudes

2.3 Genotüübiandmed

Kolmandana kasutati andmeid genotüüpide kohta. Peatükis 1.2.2 leiti viis SNP-d, mille korral statiiniravi mõjub erinevate genotüüpidega inimestele erinevalt. Broad Institute lehel on võimalik kontrollida SNP-de paariviisilist ahelduvust (*pair-wise linkage disequilibrium, pair-wise LD*) [23]. Seosekordaja $r^2 = 1$ korral on SNP-d täielikult aheldunud ja $r^2 = 0$ korral ei ole üldse. r^2 oli eelnevalt leitud kolme SNP (rs4149081, rs4149056, rs4363657) paariviisilisel võrdlemisel suuremvõrdne kui 0.887. Võrdlemaks statiinide tarvitamise seost geneetiliste variantidega, võib kasutada üht neist kolmest SNP-st, sest kui ühes aheldunud SNP-s on indiviidil riskialleel, siis on tal tõenäoliselt riskialleel ka selles teises aheldunud SNP-s. Kolmest aheldunud SNP-st olid kättesaadavad rs4149056 ja rs4363657 genotüübid, kusjuures esimese kohta olid vaid imputeeritud genotüübiandmed. Analüüsis otsustati seega kasutada rs4363657 genotüüpe. Ülejäänud kaks SNP-d (rs3846662 geenis *HMGCR*, rs2231142 geenis *ABCG2*) ei olnud aheldunud teiste leitud SNP-ga.

Eelneva taustal eraldati analüüsiks kolm genotüübitunnust: *rs4363657*, *rs3846662*, *rs2231142*. Tunnustel oli kolm väärtust: 0, 1, 2. Need arvud näitasid, mitu riskialleeli inimesel alleelipaaris oli. Tabelis 2.6 on toodud kolme SNP metsiktüüpi alleelid ja

riskialleelid. Peatükis 1.2.2 rs2231142 mõju kajastanud artikli autorid kasutasid teistpidi sekveneeritud DNA ahelat kui geenivaramu. Artiklis oli rs2231142-s riskialleeliks A ja metsiktüüpi alleeliks C. Geenivaramu andmestikus oli aga riskialleeliks A-ga komplementaarne alleel T ning metsiktüüpi alleeliks C-ga komplementaarne alleel G.

Tabel 2.6 Kasutatavate SNP-de metsiktüüpi alleelid ja riskialleelid

SNP	rs4363657	rs3846662	rs2231142
metsiktüüpi alleel	T	A	G
riskialleel	C	G	T

Tabeli 2.6 põhjal vastab tunnuse *rs4363657* väärtus 0 genotüübile TT, väärtus 1 genotüübile CT ja väärtus 2 genotüübile CC. Kõigi kolme SNP genotüübid koos sagedustega valimis on toodud tabelis 2.7.

Tabel 2.7 Genotüüpide jaotus valimis

SNP	genotüüp	isikute arv
rs4363657	TT	630
	CT	413
	CC	76
rs3846662	AA	331
	GA	513
	GG	275
rs2231142	GG	925
	TG	184
	TT	10

Rs2231142 kohta olid saadaval imputeeritud genotüübiandmed – 1119 inimese seas oli 6 inimest, kellel tunnuse *rs2231142* väärtus oli komakohaga ehk ei ennustatud päris kindlat genotüüpi. Kahel inimesel oli prognoos väiksem kui 0.5 – nemad loeti tabelis genotüübiga GG inimesteks. Neljal inimesel oli prognoos vahemikus (0.5, 1) – nemad loeti tabelis

genotüübiga TG inimesteks. Tabelist 2.7 on näha, et rs4363657-s on genotüübiga CC inimesi vaid 76 ning rs2231142-s genotüübiga TT inimesi ainult 10. Analüüsimisel kasutatakse edaspidi uusi indikaatortunnuseid $I_{rs4363657>0}$ ja $I_{rs2231142>0.5}$, mis võtavad kokku ühe ja kahe riskialleeliga inimesed. Rs3846662 genotüübid jaotusid üsna ühtlaselt, tunnus võetakse analüüsis arvesse pidevana.

3 Analüüs

3.1 Analüüsimeetodite rakendamine tarkvaras R

Leidmaks seost statiinide tarvitamise ja geneetiliste variantide vahel, kasutatakse töös ühe meetodina elukestusanalüüsi. Coxi võrdeliste riskide analüüs viiakse programmis R läbi paketi *'survival'* abil. Selles paketis on funktsioon *coxph(formula, data, ...)* [18]. Etteantava valemi üks võimalik kuju on $Surv(time, status) \sim x_1 + \dots + x_p$, kus \sim operaatorist vasakul on elukestus ja sündmuse toimumist kirjeldav binaarne tunnus ning paremal argumenttunnused. Teise meetodina kasutatakse logistilist regressiooni, mis viiakse programmis R läbi paketi *'stats'* abil, kus on funktsioon *glm(formula, data, family, ...)* [24]. Etteantavas valemis \sim operaatorist vasakul on sündmuse toimumist kirjeldav binaarne tunnus ja paremal argumenttunnused. Samuti tuleb logistilise regressiooni korral lisada tingimus *family="binomial"*.

Kõikidesse tehtavatesse mudelitesse pannakse lisaks genotüübitunnustele ka tausttunnused *vanus_statiinid* ja *bmi*. Mudelite korral on olulisuse nivoo $\alpha = 0.05$.

3.2 Andmete jagamine gruppideks

Käesolevas töös defineeritakse huvipakkuv sündmus kahel erineval viisil ning vaadatakse, milliseid tulemusi statiinide tarvitamise ja genotüüpide vahel saadakse. Peatükis 3.3 käsitletakse uuritava sündmusena statiiniravi katkestamist. Peatükis 3.4 on huvipakkuvaks sündmuseks lisaks ravi katkestamisele ka statiini vahetus. Need kaks sündmust võivad viidata sellele, et inimesele ravim ei sobinud (tekkisid kõrvaltoimed, ravim ei mõjunud vms), kuna üldjuhul jäädakse statiine võtma elu lõpuni.

Lisaks sündmuse kahel erineval viisil defineerimisele uuritakse statiinide tarvitamise seost geneetiliste variantidega nii kogu valimis kui ka alamgruppides. Tabelis 3.1 on näha, kuidas on kogu valimis kolmes SNP-s genotüübid jaotunud ning kui palju on ravi katkestajaid (sh esimese aasta jooksul katkestajaid) ning statiini vahetajaid või ravi katkestajaid (sh esimese aasta jooksul statiini vahetajaid või ravi katkestajaid) genotüüpide kaupa. Coxi võrdeliste riskide mudelitega hinnatakse, mis mõjutab ravi

katkestamise riski ning statiini vahetamise või ravi katkestamise riski. Logistilised regressioonimudelid sobivad hindama, mis mõjutab riski ravi esimese aasta jooksul katkestada ning riski esimese aasta jooksul statiini vahetada või ravi katkestada.

Tabel 3.1 Genotüüpide jaotus kogu valimis

	kogu valim (1119 inimest)								
	rs4363657			rs3846662			rs2231142		
	TT	CT	CC	AA	GA	GG	GG	TG	TT
kokku isikuid	630	413	76	331	513	275	925	184	10
katkestajaid	271	194	35	151	226	123	407	87	6
1. aasta jooksul katkestajaid	122	88	22	72	97	63	194	38	0
katkestajaid või vahetajaid	338	219	39	181	270	145	486	103	7
1. aasta jooksul katkestajaid või vahetajaid	149	106	22	85	118	74	231	46	0

Kuna peatükis 1.2.2 kajastatud SNP-sid seostati kindlate statiinidega, siis jaotatakse 1119 inimest kolme alamgruppi nii, et igas grupis on isikud, kes on võtnud statiini, mis on seotud või aheldunud ühega kolmest uuritavast SNP-st: rosuva-, simva- ja pravastatiini tarvitajad (alamgrupp 1), atorvastatiini tarvitajad (alamgrupp 2), simva- ja fluvastatiini tarvitajad (alamgrupp 3). Seejärel uuritakse iga SNP mõju vaid sellega seostatud statiini tarvitajate grupis. Tabelis 3.2 on näha kolme alamgrupi arvukus, millise SNP-ga (või sellega aheldunud SNP-ga) on neid statiine seostatud, milline on genotüüpide jaotus selles SNP-s ning kui palju on ravi katkestajaid ning statiini vahetajaid või ravi katkestajaid genotüüpide kaupa.

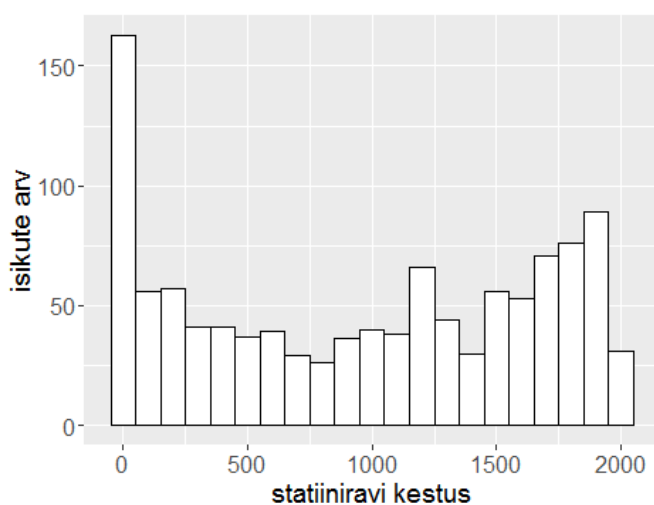
Tabel 3.2 Genotüüpide jaotus alamgruppides

	alamgrupp 1 (782 inimest)			alamgrupp 1 (439 inimest)			alamgrupp 3 (398 inimest)		
	rs4363657			rs3846662			rs2231142		
	TT	CT	CC	AA	GA	GG	GG	TG	TT
kokku isikuid	435	302	45	139	198	102	324	72	2
katkestajaid	181	153	20	58	91	42	149	37	2
katkestajaid või vahetajaid	246	178	24	82	113	57	200	47	2

Üks isik võis olla ühes, kahes või isegi kõigis kolmes alamgrupis. Näiteks kuulus igasse gruppi isik, kes tarvitas vähemalt simva- ja atorvastatiini. Tabelist 3.2 on näha, et rosuva-, simva- ja pravastatiini tarvitajate puhul on rs4363657-s genotüübiga CC inimesi vaid 45. Nagu peatükis 2.3 mainiti, kasutatakse analüüsis edaspidi uut indikaatortunnust $I_{rs4363657>0}$, mis võtab kokku CT ja CC genotüübiga inimesed. Rs3846662 puhul jaotusid genotüübid üsna ühtlaselt ka alamgrupis. Ühel isikul alamgrupis 3 oli rs2231142-s riskialleeli sageduse prognoos 0.945, tabelis 3.2 loeti ta genotüübiga TG inimeste hulka. Mudelite tegemiseks kasutatakse nagunii indikaatortunnust $I_{rs2231142>0.5}$, mis loeb selle isiku ühe või kahe riskialleeliga inimeste hulka.

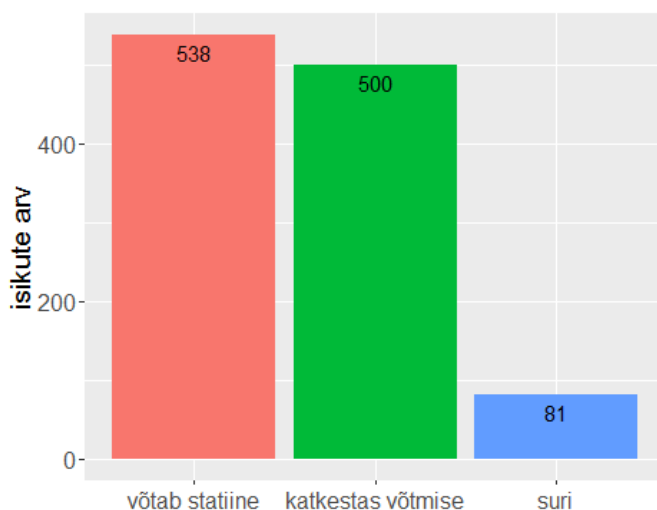
3.3 Aeg ravi katkestamiseni

Vaadates huvipakkuva sündmusena ravi katkestamist, defineeriti ravi kestust iseloomustav uus tunnus *aeg* järgnevalt: $aeg =$ viimase realiseeritud retsepti kuupäev – esimese realiseeritud retsepti kuupäev + 30. Konstant liideti ajale juurde seetõttu, et statiiniravi ei lõppe viimase retsepti ostuga, vaid kestab kuni viimase retsepti tabletid on võetud. Andmeid vaadates jäi silma, et suuremal osal retseptidest oli kogus 30 tk, mistõttu osutus 30 juurde liidetavaks päevade arvuks. Keskmiselt tarvitati statiine 967 ($SD = 683$) päeva. Statiinide tarvitamise aja jaotus päevades on toodud joonisel 3.3.



Joonis 3.3 Statiiniravi aja jaotus (päevades)

Isikuid, kes ei realiseerinud kuue kuu jooksul enne jälgimisaja lõppu ühtegi statiiniretsepti, oli 500 (vt joonis 3.4). Neid nimetakse statiiniravi katkestajateks. Kui isik suri, aga tarvitas statiine veel vähem kui kuus kuud enne surma, siis teda katkestajaks ei loetud. Selliseid isikuid oli 81. Ülejäänud 49 surnut katkestas statiinide võtmise rohkem kui kuus kuud enne surma ning neid loeti katkestajateks. Viimasena nimetati statiini võtjaks isik, kes ei olnud surnud ega ravimi võtmist katkestanud ehk oli elus ning realiseeris viimase retsepti pärast 4. veebruari 2015.



Joonis 3.4 Statiinide tarvitamise staatus

Nii elukestusanalüüsis kui ka logistilises regressioonianalüüsis läheb vaja binaarset tunnust, mis näitab huvipakkuva sündmuse toimumist (uus tunnus *pohjus_kood*). Selles alapeatükis on huvipakkuvaks sündmuseks ravi poolelijätmine. Seega ravi katkestajad said uue loodud tunnuse *pohjus_kood* väärtuseks 1. Surnud isikud ja ravimit endiselt tarvitavad isikud olid tsenseeritud ning *pohjus_kood* väärtuseks sai 0.

3.3.1 Elukestusanalüüs

Lisas 1 on toodud tarkvara R kood Coxi võrdeliste riskide mudeli hindamiseks ja riskide võrdelisuse eelduse kontrolli läbiviimiseks. Näitena on kasutatud kogu valimit, vaadates huvipakkuva sündmusena ravi katkestamist.

Kõigepealt kontrolliti, kas mõne SNP või tausttunnuse mõju katkestamise riskile tuleb statistiliselt oluline terves valimis. Statistiliselt oluliseks tuli indikaatortunnus $I_{rs4363657>0}$. Kõigi statiinide tarvitajatel, kelle on rs4363657-s genotüüp CC või CT, on 1.22 korda (22%) suurem risk katkestada statiiniravi võrreldes isikutega, kellel on genotüüp TT ($p = 0.025$). Lisas 2 on toodud elukestusfunktsiooni Kaplan-Meieri hinnang rs4363657 genotüüpide kaupa ning elukestuse mediaan rs4363657 genotüüpide kaupa.

Teisena vaadati, kas indikaatortunnuse $I_{rs4363657>0}$ mõju võiks tulla olulisem nende isikute seas, kes on tarvitanud selle SNP-ga seostatud statiini ehk rosuva-, simva või pravastatiini. Samuti lisati mudelisse tausttunnused. Statistiliselt olulisteks tulid tunnused $I_{rs4363657>0}$ ja *vanus_statiinid*. Rosuva-, simva- ja pravastatiini tarvitajatel, kellel on rs4363657-s genotüüp CC või CT, on 1.41 korda (41%) suurem risk katkestada statiiniravi kui isikutel, kellel on rs4363657-s genotüüp TT ($p = 0.001$), eeldusel et vanus statiiniravi alustades on sama. Samade statiinide tarvitajate seas on isikul, kes alustas statiiniravi üks aasta vanemana kui teine isik, keskmiselt $1 - 0.99 = 0.01$ korda (1%) väiksem risk katkestada statiiniravi ($p = 0.022$), eeldusel et isikutel on rs4363657-s sama genotüüp.

Alamgruppides 1 ja 2 ei mõjuta katkestamise riski SNP-d ega tausttunnused. Kokkuvõtte kõigist Coxi võrdeliste riskide mudelitest statiiniravi katkestamisele on toodud tabelis 3.5.

Tabel 3.5 Coxi võrdeliste riskide mudelite tulemused katkestamise korral

	tunnus	$\hat{\beta}$ (SE)	p-väärtus	$e^{\hat{\beta}}$	$e^{\hat{\beta}}$ 95%-UI
kogu valimis	$I_{rs4363657>0} = 1$	0.203 (0.090)	0.025	1.225	(1.026, 1.462)
	<i>rs3846662</i>	-0.014 (0.062)	0.819	0.986	(0.874, 1.113)
	$I_{rs2231142>0.5} = 1$	0.0889 (0.115)	0.441	1.093	(0.872, 1.370)
	<i>vanus_statiinid</i>	-0.007 (0.004)	0.058	0.993	(0.986, 1.000)
	<i>bmi</i>	-0.018 (0.010)	0.058	0.982	(0.964, 1.001)
alamgrupis 1	$I_{rs4363657>0} = 1$	0.340 (0.107)	0.001	1.405	(1.139, 1.734)
	<i>vanus_statiinid</i>	-0.010 (0.005)	0.022	0.990	(0.981, 0.999)
	<i>bmi</i>	-0.015 (0.011)	0.180	0.985	(0.964, 1.007)
alamgrupis 2	<i>rs3846662</i>	0.002 (0.100)	0.988	1.001	(0.824, 1.218)
	<i>vanus_statiinid</i>	0.003 (0.006)	0.599	1.003	(0.992, 1.015)
	<i>bmi</i>	-0.029 (0.015)	0.055	0.971	(0.942, 1.001)
alamgrupis 3	$I_{rs2231142>0.5} = 1$	0.154 (0.182)	0.396	1.167	(0.817, 1.666)
	<i>vanus_statiinid</i>	-0.008 (0.006)	0.226	0.992	(0.980, 1.005)
	<i>bmi</i>	-0.020 (0.016)	0.207	0.980	(0.950, 1.011)

3.3.2 Logistiline regressioonanalüüs

Järgnevalt otsiti esimese aasta jooksul ravi katkestamise šanssi mõjutavaid tunnuseid kõigi 1119 inimese korral. Lisas 3 on toodud tarkvara R kood logistilise regressioonmudeli hindamiseks. Näitena on kasutatud kogu valimit, vaadates huvipakkuva sündmusena ravi katkestamist. Statistiliselt oluliseks osutus siin vaid tausttunnus *vanus_statiinid*. Kui vaadata kaht isikut, kellest üks on statiiniravi alustanud

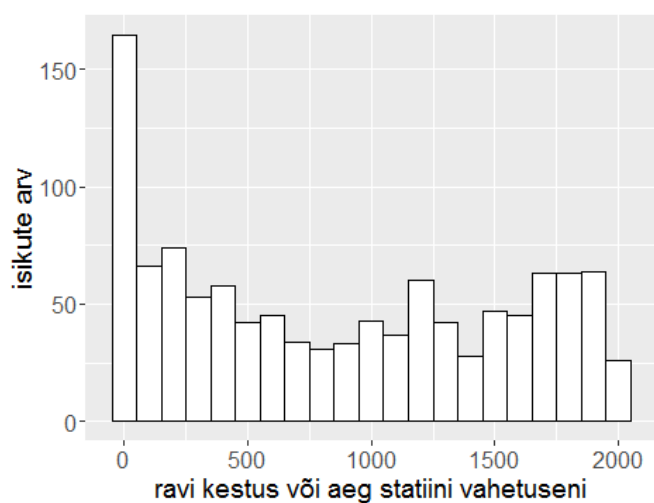
üks aasta vanemana, siis tema šanss esimese aasta jooksul ravi katkestada on $1 - 0.99 = 0.01$ korda (1%) väiksem ($p = 0.023$). Kokkuvõtte logistilise regressioonimudeli tulemustest statiiniravi katkestamisele on tabelis 3.6.

Tabel 3.6 Logistilise regressioonimudeli tulemused katkestamise korral

tunnus	$\hat{\beta}$ (SE)	p-väärtus	$e^{\hat{\beta}}$	$e^{\hat{\beta}}$ 95%-UI
$I_{rs4363657>0} = 1$	0.211 (0.149)	0.156	1.235	(0.922, 1.653)
$rs3846662$	0.032 (0.101)	0.749	1.033	(0.847, 1.259)
$I_{rs2231142>0.5} = 1$	-0.089 (0.199)	0.653	0.915	(0.612, 1.338)
$vanus_statiinid$	-0.014 (0.006)	0.023	0.986	(0.975, 0.998)
bmi	-0.010 (0.015)	0.524	0.990	(0.961, 1.020)

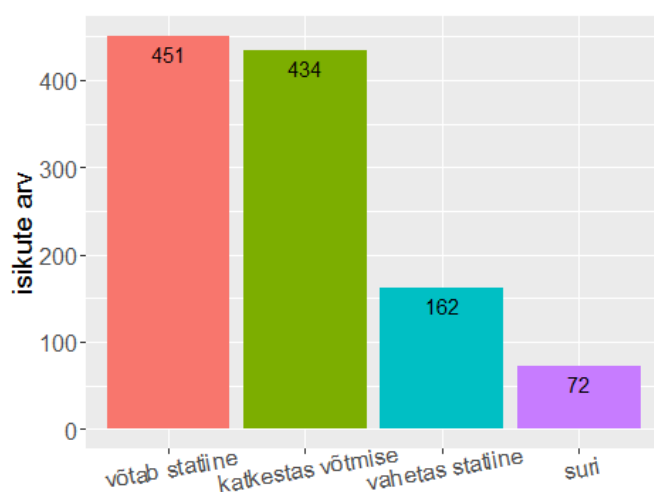
3.4 Aeg ravi katkestamiseni või statiini vahetuseni

Kasutades digiretseptide andmeid, ilmnes, et 162 isikut vahetas uuringuperioodi jooksul statiiniravimit. Neid nimetatakse statiiniravimi vahetajateks, mis tähendab, et nad võtsid mitut erinevat statiini, näiteks alguses simva- ja seejärel rosuvastatiini. 155 inimest tarvitas kaht ning seitse inimest kolme erinevat statiini. Võttes huvipakkuvaks sündmuseks ka ravimi vahetuse, tuli selliste isikute puhul tunnuse *aeg* väärtus uuesti arvutada. See defineeriti järgnevalt: $aeg =$ esimene teise unikaalse statiini realiseerimiskuupäev – esimene esimese unikaalse statiini realiseerimiskuupäev + 30. Konstant 30 liideti ajale juurde jällegi seetõttu, et statiiniravi ei lõppe viimase retsepti ostuga, vaid kestab kuni viimase retsepti tabletid on võetud. Keskmise aeg ravi lõpetamiseni või statiini vahetuseni oli 872 ($SD = 666$) päeva. Statiinide katkestamise või vahetamise aja jaotus päevades on toodud joonisel 3.7.



Joonis 3.7 Statiini vahetamise või katkestamise aja jaotus (päevades)

Isikuid, kes ei olnud statiine vahetanud ega kuus kuud enne jälgimisaja lõppu statiiniretsepti realiseerinud, oli 434 (vt joonis 3.8). Neid nimetatakse statiiniravi katkestajateks. Kui isik suri, aga ei olnud statiini vahetanud ning tarvitas neid veel vähem kui kuus kuud enne surma, siis teda vahetajaks ega katkestajaks ei loetud. Selliseid isikuid oli 72. Ülejäänud 58 surnut vahetas uuringuperioodi jooksul statiini või katkestas nende võtmise rohkem kui kuus kuud enne surma. Viimasena loeti statiini võtjaks isik, kes oli elus, võttis üht statiini kogu uuringu jooksul ning realiseeris viimase retsepti pärast 4. veebruari 2015.



Joonis 3.8 Statiinide tarvitamise staatus

Kuna selles alapeatükis on huvipakkuvaks sündmuseks ravimivahetus või ravi pooleljätmine, said statiini vahetajad ja katkestajad uue tunnuse *pohjus_kood* väärtuseks

1. Surnud isikud ja ravimit endiselt tarvitavad isikud olid tsenseeritud ning *pohjus_kood* väärtuseks sai 0.

3.4.1 Elukestusanalüüs

Esmalt vaadati, kas mõne SNP või tausttunnuse mõju katkestamise või vahetamise riski tuleb statistiliselt oluline terves valimis. Statistiliselt olulisteks tulid tunnused *vanus_statiinid* ja *bmi*. Isikul, kes alustas statiiniravi üks aasta vanemana kui teine isik, on $1 - 0.99 = 0.01$ korda (1%) väiksem risk katkestada statiiniravi või vahetada statiini ($p = 0.033$), eeldusel et isikutel on sama kehamassiindeks. Kui vaadata kaht isikut, kellest ühe kehamassiindeks on ühe ühiku võrra suurem, siis tema risk katkestada ravi või vahetada statiini on $1 - 0.98 = 0.02$ korda (2%) väiksem ($p = 0.045$).

Teisena tehti jällegi mudelid alamgruppides. Statistiliselt oluliseks tuli alamgrupis 1 tausttunnus *vanus_statiinid*. Rosuva-, simva- ja pravastatiini tarvitajate seas on isikul, kes alustas statiiniravi üks aasta vanemana kui teine isik, $1 - 0.99 = 0.01$ korda (1%) väiksem risk katkestada ravi või vahetada statiini ($p = 0.005$). Võttes nii kogu valimis kui ka alamgrupis 1 huvipakkuvaks sündmuseks ainult statiiniravi katkestamise, tuli rs4363657 erinevate genotüüpide korral katkestamise risk erinev, võttes aga huvipakkuvaks sündmuseks statiiniravi katkestamise või statiini vahetamise, ei tulnud risk erinev. See tähendab, et geenidoonorid pigem lõpetavad statiinide võtmise, kui vahetavad retsepti. Tõenäoliselt ei teata, et mittesobivuse põhjus võib olla konkreetnes statiinis, mitte üleüldiselt statiinides.

Alamgrupis 2 tuli statistiliselt oluliseks kehamassiindeks. Atorvastatiini tarvitajate seas väheneb ühe ühiku kehamassiindeksi suurenedes ravi katkestamise või statiini vahetamise risk $1 - 0.97 = 0.03$ korda (3%) ($p = 0.042$). Alamgrupis 3 tuli statistiliselt oluliseks tunnus *vanus_statiinid*. Simva- või fluvastatiini tarvitajal, kes alustas statiiniravi üks aasta vanemana kui teine isik, on $1 - 0.99 = 0.01$ korda (1%) väiksem risk katkestada statiiniravi või vahetada statiini ($p = 0.016$).

Ükski SNP oluline ei tulnud, olulised tulid ainult tausttunnused. Kokkuvõtte kõigist Coxi võrdeliste riskide mudelitest statiiniravi katkestamisele või statiini vahetamisele on toodud tabelis 3.9.

Tabel 3.9 Coxi võrdeliste riskide mudelite tulemused katkestamise ja vahetamise korral

	tunnus	$\hat{\beta}$ (SE)	p- väärtus	$e^{\hat{\beta}}$	$e^{\hat{\beta}}$ 95%-UI
kogu valimis	$I_{rs4363657>0} = 1$	0.077 (0.083)	0.355	1.080	(0.918, 1.271)
	<i>rs3846662</i>	-0.025 (0.057)	0.660	0.975	(0.873, 1.090)
	$I_{rs2231142>0.5} = 1$	0.059 (0.106)	0.576	1.061	(0.862, 1.310)
	<i>vanus_statiinid</i>	-0.007 (0.003)	0.033	0.993	(0.986, 0.999)
	<i>bmi</i>	-0.017 (0.009)	0.045	0.983	(0.966, 0.999)
alamgrupp 1	$I_{rs4363657>0} = 1$	0.158 (0.096)	0.098	1.172	(0.971, 1.413)
	<i>vanus_statiinid</i>	-0.011 (0.004)	0.005	0.989	(0.981, 0.997)
	<i>bmi</i>	-0.014 (0.010)	0.149	0.986	(0.967, 1.005)
alamgrupp 2	<i>rs3846662</i>	-0.022 (0.086)	0.802	0.979	(0.826, 1.159)
	<i>vanus_statiinid</i>	0.004 (0.005)	0.484	1.004	(0.994, 1.014)
	<i>bmi</i>	-0.027 (0.013)	0.042	0.974	(0.949, 0.999)
alamgrupp 3	$I_{rs2231142>0.5} = 1$	0.047 (0.161)	0.771	1.048	(0.764, 1.437)
	<i>vanus_statiinid</i>	-0.014 (0.006)	0.016	0.986	(0.975, 0.997)
	<i>bmi</i>	-0.018 (0.014)	0.192	0.983	(0.957, 1.009)

3.4.2 Logilistine regressioonianalüüs

Esimese aasta jooksul ravi katkestamise või statiini vahetamise šanssi mõjutavaid tunnuseid hakatati samuti leidma kõigi 1119 inimese pealt. Statistiliselt oluliseks tuli jällegi tausttunnus *vanus_statiinid*. Kui vaadata kaht isikut, kellest üks on statiiniravi

alustanud üks aasta vanemana, siis tema šanss katkestada ravi või vahetada statiini on $1 - e^{-0.018} \approx 1 - 0.98 = 0.02$ korda (2%) väiksem ($p = 0.001$). Kokkuvõtte logistilise regressioonimudeli tulemustest ravi katkestamisele või statiini vahetamisele on tabelis 3.10.

Tabel 3.10 Logistilise regressioonimudeli tulemused katkestamise või vahetamise korral

tunnus	$\hat{\beta}$ (SE)	p-väärtus	$e^{\hat{\beta}}$	$e^{\hat{\beta}}$ 95%-UI
$I_{rs4363657>0} = 1$	0.162 (0.140)	0.247	1.176	(0.893, 1.549)
$rs3846662$	0.029 (0.095)	0.758	1.029	(0.854, 1.241)
$I_{rs2231142>0.5} = 1$	-0.072 (0.186)	0.698	0.931	(0.640, 1.331)
$vanus_statiinid$	-0.018 (0.006)	0.001	0.982	(0.972, 0.993)
bmi	-0.015 (0.014)	0.282	0.985	(0.957, 1.012)

Kokkuvõte

Bakalaureusetöö eesmärk oli Tartu Ülikooli Eesti Geenivaramu geenidoonorite põhjal uurida, kas invidiidid, kellel on kõrvaltoimetega seondatud genotüüp, jätavad statiinide tarvitamise keskmisest sagedamini pooleli. Teisalt vaadati, kas vanus statiiniravi alustades ja kehamassiindeks mõjutavad ravimi tarvitamise katkestamist. Nende hindamiseks kasutati Tartu Ülikooli Eesti Geenivaramu ja Eesti Haigekassa digiretseptide andmeid. Vaadeldav periood oli 4. jaanuar 2010 kuni 4. august 2015, mille jooksul realiseerus 1119 geenidoonori seas kokku 12 999 statiiniretsepti. Analüüsimiseks tehti Coxi võrdeliste riskide ja logistilise regressiooni mudelid.

Osutus, et rosuva-, simva- ja pravastatiini võtjatel, kellel on SNP-s rs4363657 genotüüp CC või CT, on 41% suurem risk katkestada statiiniravi kui isikutel, kellel on genotüüp TT. Teine oluline tulemus ilmnis, kui vaadata lisaks katkestamisele huvipakkuva sündmusena ka statiini vahetamist, mis võib samuti viidata kõrvaltoimete esinemisele. Sellisel juhul ei tulnud ravi katkestamise või statiini vahetamise risk erinevate genotüüpidega invidiididel erinev. Sellest võib järeldada, et kõrvalmõjusid kogevad geenidoonorid pigem lõpetavad statiinide võtmise, kui vahetavad retsepti. Tõenäoliselt ei teata, et mitesobivuse põhjus võib olla konkreetnes statiinis, mitte üleüldiselt statiinides.

Kehamassiindeks ja vanus statiiniravi alustades mõjutavad katkestamise või vahetamise riski. Näiteks kõigi statiinide tarvitajate seas on ühe kg/m² suurema kehamassiindeksiga isikul 2% väiksem risk ravi katkestada või statiini vahetada. Üks aasta vanemana statiiniravi alustanud on 1% väiksem risk ravi katkestada või statiini vahetada. Vanus statiiniravi alustades avaldab ka mõju esimese aasta jooksul ravi katkestamise või statiini vahetamise šanssi. Üks aasta vanemana statiiniravi alustanud isikul on šanss esimese aasta jooksul ravi katkestada või statiini vahetada 2% väiksem.

Kokkuvõttes võib öelda, et patsientidel, kellele soovitaks määrata ravi statiinidega, peaks lisaks fenotüübitunnuste arvesse võtmisele kasutama ka geneetilist testimist. Tööd on võimalik edasi arendada, vaadates lisaks statiinidele ka teiste ravimite korral genotüüpide mõju katkestamise või vahetamise riskile. Samuti oli käesolevas andmestikis valimimaht mõnedes gruppides liiga väike. Näiteks simva- ja fluvastatiini tarvitajate seas oli SNP-s

rs2231142 genotüübiga TT inimesi vaid 2. Mõlemad inimesed on ravi katkestanud, mis võib viidata kõrvaltoimetele, kuid valim oli niivõrd väike, et järeldusi selle põhjal teha ei saanud.

Viited

[1] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL (vaadatud 25.04.2016)

<https://www.R-project.org/>

[2] A. Heinaru (2012). *Geneetika*. Tartu: Tartu Ülikooli Kirjastus

[3] The International Hapmap Consortium (2003). *The International HapMap Project*. Nature 426, 789-796 (vaadatud 27.04.2016)

Kättesaadav: <http://www.nature.com/nature/journal/v426/n6968/full/nature02168.html>

[4] B. Carlson (2008). *SNPs - A shortcut to personalized medicine*. Biomarket Trends, Vol. 28, No. 12 (vaadatud 27.04.2016)

Kättesaadav: <http://www.genengnews.com/gen-articles/snps-a-shortcut-to-personalized-medicine/2507/>

[5] Toimetaja A. V. Mikelsaar (2010). *Pärilikkusmeditsiin*. Medicina kirjastus, Tallinn

[6] Kolesterool. kliinikum.ee, URL (vaadatud 25.04.2016)

<http://www.kliinikum.ee/yhendlabor/images/stories/kasiraamat/K/kolesterool.pdf>

[7] D. M. Roden (2016). *Cardiovascular pharmacogenomics: current status and future directions*. Journal of Human Genetics 61, 79-85 (vaadatud 27.04.2016)

Kättesaadav: <http://www.nature.com/jhg/journal/v61/n1/pdf/jhg201578a.pdf>

[8] R. Valiyil, L. Christopher-Stine (2010). *Drug-related myopathies of which the clinician should be aware*. Current Rheumatology Reports, Vol. 12, Issue 3, 213-220 (vaadatud 27.04.2016)

Kokkuvõte kättesaadav: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3092639/>

[9] H. Kahn, H. M. Loit (2009). *Tervise ABC*. Tallinn: Valgus

[10] E. Link, S. Parish, J. Armitage et al. (2008). *SLCO1B1 variants and statin-induced myopathy--a genomewide study*. N. Engl. J. Med. 359, 789–799 (vaadatud 27.04.2016)

Kokkuvõte kättesaadav: <http://www.ncbi.nlm.nih.gov/pubmed/18650507>

[11] J. E. Keskitalo, M. K. Pasanen, P. J. Neuvonen, M. Niemi (2009). *Different effects of the ABCG2 c.421C>A SNP on the pharmacokinetics of fluvastatin, pravastatin and simvastatin*. Pharmacogenomics 10, 1617-1624 (vaadatud 27.04.2016)

Kokkuvõte kättesaadav: <http://www.ncbi.nlm.nih.gov/pubmed/19842935>

[12] J. Y. Chung, S. K. Cho, E. S. Oh et al. (2012). *Effect of HMGCR variant alleles on low-density lipoprotein cholesterol-lowering response to atorvastatin in healthy Korean subjects*. J Clin Pharmacol. 52, 339-346 (vaadatud 27.04.2016)

Kokkuvõte kättesaadav: <http://www.ncbi.nlm.nih.gov/pubmed/21427285>

[13] H. Akao, E. Polisecki, K. Kajinami et al. (2012). *Genetic variation at the SLCO1B1 gene locus and low density lipoprotein cholesterol lowering response to pravastatin in the elderly*. Atherosclerosis 220, 413-417 (vaadatud 27.04.2016)

Kokkuvõte kättesaadav: <https://www.ncbi.nlm.nih.gov/pubmed/22189199>

[14] M. Hu, V. W. Mak, B. Tomlinson (2012). *Intronic variants in SLCO1B1 related to statin-induced myopathy are associated with the low-density lipoprotein cholesterol response to statins in Chinese patients with hyperlipidaemia*. Pharmacogenet.Genomics 22, 803–806 (vaadatud 27.04.2016)

Kokkuvõte kättesaadav: <http://www.ncbi.nlm.nih.gov/pubmed/22668755>

[15] A. Irs (2015). *Kardiovaskulaarravimite kasutamise suundumused. 20 aastat ravimistatistikat Eestis*, 40-55 (vaadatud 27.04.2016)

Kättesaadav: <http://www.ravimiamet.ee/sites/default/files/Kardiovaskulaarravimite%20suundumused.pdf>

[16] D. G. Kleinbaum, M. Klein (2005). *Survival analysis: A self-learning text, second edition*

[17] K. Fischer (2007). *Elukestusanalüüs*. Loengukonspekt. Tartu: Tartu Ülikool, tervishoiu instituut

[18] S. J Walters (2009). *What is a Cox model? What if...?* series, second edition (vaadatud 27.04.2016)

Kättesaadav:

http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/cox_model.pdf

[19] Fit Proportional Hazards Regression Model. R Documentation, URL (vaadatud 25.04.2016)

<https://stat.ethz.ch/R-manual/R-devel/library/survival/html/coxph.html>

[20] Test the Proportional Hazards Assumption of a Cox Regression. R Documentation, URL (vaadatud 25.04.2016)

<https://stat.ethz.ch/R-manual/R-devel/library/survival/html/cox.zph.html>

[21] E. Käärik (2015). *Andmeanalüüs II*. Loengukonspekt. Tartu: Tartu Ülikool, matemaatilise statistika instituut (vaadatud 27.04.2016)

Kättesaadav: <http://dspace.ut.ee/bitstream/handle/10062/35401/AndmeanalüüsII.pdf>

[22] The Anatomical Therapeutic Chemical Classification System with Defined Daily Doses (ATC/DDD), WHO, URL (vaadatud 25.04.2016)

<http://www.who.int/classifications/atcddd/en/>

[23] Pairwise LD, SNP Annotation and Proxy Search, Broad Institute, URL (vaadatud 25.04.2016)

<https://www.broadinstitute.org/mpg/snap/ldsearchpw.php>

[24] Generalized linear models, Quick-R, URL (vaadatud 25.04.2016)

<http://www.statmethods.net/advstats/glm.html>

Lisad

Lisa 1. Coxi võrdeliste riskide mudeli hindamine tarkvaras R

Näide Coxi võrdeliste riskide mudeli hindamisest, vaadates huvipakkuva sündmusena statiiniravi katkestamist (kogu valimis).

```
# mudel kolme genotüübitunnusega ja kahe tausttunnusega
> cox_kat <- coxph(Surv(aeg, pohjus_kood) ~ I(rs4363657>0) + rs3846662
  + I(rs2231142>0.5) + vanus_statiinid + bmi, kokku_kat)
# mudeli parameetri hinnangute ja muude oluliste statistikute
  väljastamine
> summary(cox_kat)

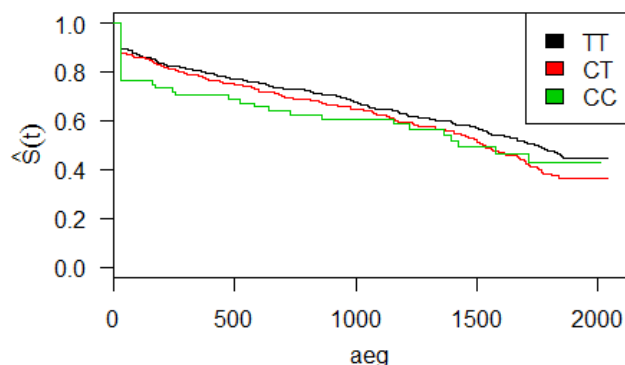
# riskide võrdelisuse eelduse kontroll
> cox.zph(cox_kat)
```

Lisa 2. Elukestusfunktsiooni hindamine Kaplan-Meieri meetodiga tarkvaras R

Näide elukestusfunktsiooni hindamisest Kaplan-Meieri meetodiga, vaadates huvipakkuva sündmusena statiiniravi katkestamist (kogu valimis). Kuna tunnus *rs4363657* mõjutab ravi katkestamise riski, siis vaadatakse, millised on erinevate genotüüpide korral elukestusfunktsiooni kõverad.

```
# elukestusfunktsiooni hinnang Kaplan-Meieri meetodiga rs4363657
genotüüpide kaupa

> surv1 <- survfit(Surv(aeg, pohjus_kood) ~ rs4363657,
                   data = kokku_kat)
> plot(surv1, conf = F, mark.time = F, col = 1:3)
> title(xlab="aeg", ylab= expression(hat(S)(t)), line=2.3)
> legend("topright", legend=c("TT", "CT", "CC"), fill=1:3)
```



Joonis lisa 2. Elukestusfunktsiooni hinnang Kaplan-Meieri meetodiga *rs4363657* genotüüpide kaupa

Jooniselt on näha, et genotüüpide CC ja CT elukestusfunktsiooni kõverad on madalamal kui genotüübi TT. Seega CC ja CT genotüüpidega indiviididel on tõenäosus mitte katkestada ravi enne mingit ajahetke t või selle ajal väiksem kui TT genotüübiga indiviididel.

Elukestuse mediaan kirjeldab, milliseks ajahetkeks pärast ravi alustamist on pooled isikutest statiinide võtmise katkestanud.

```
# elukestuse mediaani väljastamine
```

```
> surv1
```

```
          n events median 0.95LCL 0.95UCL
rs4363657=0 630   271  1739   1561     NA
rs4363657=1 413   194  1525   1406  1696
rs4363657=2  76    35  1422   1157     NA
```

Pooled isikutest, kelle on rs4363657-s genotüüp TT, on 1739 päeva jooksul pärast ravi alustamist statiinide võtmise katkestanud. Pooled isikutest, kelle on rs4363657-s genotüüp CT, on 1525 päeva jooksul pärast ravi alustamist statiinide võtmise katkestanud. Pooled isikutest, kelle on rs4363657-s genotüüp CC, on 1422 päeva jooksul pärast ravi alustamist statiinide võtmise katkestanud. Seega, mida rohkem riskialleele, seda varem pooled isikutest ravi lõpetavad.

Lisa 3. Logistilise regressioonimudeli hindamine tarkvaras R

```
# mudel kolme genotüübitunnusega ja kahe tausttunnusega
> log_kat <- glm(pohjus_kood2 ~ I(rs4363657>0) + rs3846662 +
                I(rs2231142>0.5) + vanus_statiinid + bmi,
                data = kokku_kat, family = binomial())
# mudeli parameetri hinnangute ja muude oluliste statistikute
väljastamine
> summary(log_kat)

# šansside 95%-usaldusintervall
> exp(confint(log_kat))
```

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Enelin Haviko (sünnikuupäev 09.10.1994)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Statiinide tarvitamise seos geneetiliste variantidega TÜ Eesti Geenivaramu geenidoonoritel“, mille juhendajad on Lili Milani ja Krista Fischer,
 - a. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - b. 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, **29.04.2016**