



# Vähki haigestumise riski hindamine vanuse, soo ja aasta järgi



Tervise Arengu Instituut  
National Institute for Health Development

Tervise Arengu Instituut

# **Vähki haigestumise riski hindamine vanuse, soo ja aasta järgi**

Viktoria Kirpu

Tallinn 2020

Tervise Arengu Instituudi **missioon** on luua ja jagada teadmisi, et tõendus põhisel mõjutada tervist toetavaid hoiakuid, käitumist, poliitikat ja keskkonda eesmärgiga suurendada inimeste heaolu Eestis.

Käesolev analüüs on algupäraselt valminud Tartu Ülikooli matemaatika ja statistika instituudi magistritööna.

Täna juhendajat Märt Mölsi ja kolleege Margit Mägi ja Ingrid Valdmaad nõuannete ja kommentaaride eest.

Väljaande andmete kasutamisel viidata allikale.

Soovitatav viide käesolevale väljaandele: Kirpu V. Vähki haigestumise riski hindamine vanuse, soo ja aasta järgi. Tallinn: Tervise Arengu Instituut; 2020.

# Sisukord

Kokkuvõte .....	6
1. Andmed .....	9
1.1 Vähiregister .....	9
1.1.1 Kasutatud andmed .....	9
1.1.2 Analüüsis käsitletud vähiliigid .....	10
1.2 Statistikaameti andmebaas .....	14
2. Metoodika .....	15
2.1 Mudel .....	15
2.1.1 Poissoni mudel .....	15
2.1.2 Grupeeritud andmed .....	15
2.1.3 Mudeli hindamine .....	16
2.1.4 Ala- ja ülehajuvus .....	17
2.2 Splainid .....	18
2.3 Parima mudeli valimine .....	19
2.4 Seost kirjeldava funktsiooni kasvamis-, kahanemis- ja käänukohad .....	20
2.5 Binoomtest .....	22
2.6 Haigestumuskordaja standardiseerimine vanuse ja soo järgi .....	22
3. Kirjeldav analüüs .....	25
3.1 Kõik paikmed .....	25
3.2 Emakakaelavähk .....	27
3.3 Eesnäärmevähk .....	29
4. Mudelite kirjeldus ja prognoosivõime .....	31
4.1 Kõik paikmed .....	31
4.1.1 Vanus .....	32
4.1.2 Diagnoosimise aasta .....	34
4.1.3 Mudeli prognoosivõime .....	37
4.2 Emakakaelavähk .....	38
4.2.1 Vanus .....	39
4.2.2 Diagnoosimise aasta .....	40
4.2.3 Mudeli prognoosivõime .....	42
4.3 Eesnäärmevähk .....	43
4.3.1 Vanus .....	43
4.3.2 Diagnoosimise aasta .....	45
4.3.3 Mudeli prognoosivõime .....	46
4.4 Haigestunute arvu standardiseerimine rahvastiku soo ja vanuse järgi .....	47
Kasutatud kirjandus .....	50

Lisad .....	52
Lisa 1. Pahaloomuliste kasvajate grupid ja esinemiste arv kokku 1968.-2017. aastal .....	52
Lisa 2. Andmete ettevalmistamine (R).....	54
Lisa 3. Andmete kirjeldav analüüs (R).....	57
Lisa 4. Parima mudeli valimine (R).....	61
Lisa 5. Mudeli konstrueerimine ja analüüs (R) .....	62
Lisa 6. Haigestumise riski standardiseerimine rahvas-tiku soo- ja vanusjaotuse järgi (R).....	68
Lisa 7. Prognoosi prognoosivõime uurimine (R).....	69
Lisa 8. Mudeli $s(x, y, z)$ liikme kuju kõikide paikmete korral.....	70
Lisa 9. Mudeli $s(x, y, z)$ liikme kuju emakakaelavähi korral .....	71
Lisa 10. Mudeli $s(x, y, z)$ liikme kuju eesnäärmevähi korral.....	71

## Lühikokkuvõte

### Vähki haigestumise riski hindamine vanuse, soo ja aasta järgi

Magistritöös uuriti vähki haigestumise riski patsiendi vanuse, diagnoosimise aasta ja võimalusel patsiendi soo järgi. Sissejuhatuseks toodi töös välja lühike ülevaade kasutatud vähiregistri ja Statistikaameti andmebaasi rahvaarvu andmetest, vähi olemusest ning analüüsis käsitletud vähipaikmetest: eesnääre ja emakakael. Mudelite loomisel kasutati Poissoni ja kvaasi-Poissoni regressiooni, andmetele leiti Clopper–Pearson'i usaldusintervallid ning mudeli kasvamis-, kahanemis- ja käänukohtade tuvastamiseks rakendati  $\chi^2$ -testi ja  $t$ -testi. Graafikute abil uuriti standardiseeritud haigestumuskordaja muutumist. Uuriti 2017. aasta andmete alusel loodud mudelite prognoosivõimet.

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

**Märksõnad:** Mudelid, riskianalüüs, Poissoni regressioon, kvaasi-Poissoni regressioon, splineid,  $\chi^2$ -test,  $t$ -test, Clopper–Pearson'i usaldusintervallid, standardiseerimine, simulatsioon, R (programmeerimiskeel).

## Abstract

### Assessing the risk of cancer by age, gender and year

The objective of this master's thesis is to study the risk of cancer by patient's age, year of diagnosis and, if possible, patient's gender. This study provides a brief overview of the Estonian Cancer Registry's data and population data published by Statistics Estonia. It also gives a description of the cancer in general and the cancer types covered in this thesis: prostate and cervix uteri. The models are constructed using the Poisson and quasi-Poisson regressions, Clopper–Pearson's confidence intervals are found for the data, and the  $\chi^2$ -test and  $t$ -test are used to detect model shrinking, growing and inflection points. Graphs have been used to examine changes in the standardized morbidity rate. The model's ability to predict is tested on the data of 2017.

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

**Keywords:** Models, Poisson regression, quasi-Poisson regression, splines,  $\chi^2$ -test,  $t$ -test, Clopper–Pearson's confidence intervals, standardizing, confidence intervals, simulation, R (programming language).

## Kokkuvõte

Eestis ei ole varem teostatud selliseid analüüse, kus uuritaks vähki haigestumise riski muutumist diagnoosimise aasta, patsiendi soo ja vanuse järgi ning kuhu kaasataks kõiki alates 1968. aastast vähiregistrist kättesaadavaid andmeid. Antud analüüsi valmimise ajal oli Eesti vähiregistrist võimalik kasutamiseks saada andmeid aastate 1968–2017 kohta. Lisaks kasutati haigestumisrisiki leidmiseks rahvaarvu andmeid Statistikaameti andmebaasist.

Haigestumisrisiki uuriti kõikide paikmete korral kokku naistel ja meestel eraldi ning ainult naiste puhul emakakaelavähki ja ainult meeste puhul eesnäärmevähki. Et jooniste abil uurida patsiendi vanuse mõju haigestumisrisiki jaoks, fikseeriti aasta 2000 ning et graafiku kaudu uurida diagnoosimise aasta mõju riski jaoks, fikseeriti vanus 60.

Kõiki paikmeid uurides saadi teada, et kõrgemas vanuses haigestuvad mehed ligi kaks korda rohkem vähki kui naised ja noortel on tunduvalt väiksem risk haigestuda vähki kui eakatel. Lisaks on meeste puhul haigestumisrisk aastatel 1968–2016 natuke rohkem kasvanud kui naistel. Emakakaelavähi korral kasvab risk haigestuda juba alates 20ndatest eluaastatest ja veidi väheneb kõrgemas eas. Kõige kõrgem on naistel risk emakakaelavähki haigestuda vanuses 50–80. Aastate jooksul on risk haigestuda emakakaelavähki olnud pidevas muutumises ja mingeid otseseid trende välja tuua pole võimalik. Eesnäärmevähi puhul on risk haigestuda väga järsult kasvanud aastatel 2000–2010. Samuti hakkab risk haigestuda eesnäärmevähki kasvama alates meeste 40ndatest eluaastatest ja on eriti kõrge vanuses 70–80.

Uuringu käigus saadi teada, et mitte kõikide vähipaikmete korral pole võimalik luua reaalsust peegeldavaid mudeleid. Mudeleid on mõttekas luua ainult nende vähipaikmete korral, mille puhul on võimalikult palju esinemise juhte, et mudelid ei oleks tugevalt mõjutatud üksikjuhtumitest. Väga hästi õnnestus luua mudeleid kõikide paikmete kokku ning eesnäärmevähi korral, mida saaks rakendada ka praktikas. Loodud emakakaelavähi mudel oli ebastabiilsem, sest selle vähipaikme korral esineb ka vähe haigestumise juhte. Antud mudel on tugevalt mõjutatud üksikjuhtumitest. Uurides mudeli prognoosivõimet 2017. aastaks saadi teada, et loodud mudelite põhjal on lühiajaliselt võimalik üsna hästi prognoosida haigestunute arvu ja seda kõikide uuritud paikmete korral.

Läbi viidud analüüsi põhjal sai uurija teada, et kui aastatel 1968–2015 oleks rahvastiku vanuseline ja sooline jaotus olnud samasugune nagu aastal 2016, siis keskmiselt oleks vähki haigestunuid olnud palju rohkem. Peale haigestumiste arvu standardiseerimist sai uurija teha ka järelduse, et haigestumine on aastatega statistiliselt olulisel määral kasvanud.

Seost kirjeldava funktsiooni kasvamis-, kahanemis- ja käänukohtade identifitseerimiseks kasutati  $\chi^2$ -testi või  $t$ -testi. See lubas mudeli põhjal teha kindlaks iga vanuse ja diagnoosimise aasta korral eraldi kasvamis- ja kahanemiskiirkonnad. Antud uuringus leitud mudeli kasvamis- ja kahanemised on kogutud kõik ühte tabelisse (vt tabelit 1).

**Tabel 1.** Mudelite kasvamis- (↑) ja kahanemiskohtade (↓) kokkuvõte

Vähi paige	Vanus		Aasta	
	Mees	Naine	Mees	Naine
<b>Kõik</b>	↓ 2-8, ↑ 10-78, ↓ 81-85+	↓ 3-8, ↑ 10-78, ↓ 82-85+	↓ 1968-1969, ↑ 1971-1973, ↑ 1980-1997, ↑ 2002-2011	↑ 1982-1987, ↑ 1991-1998, ↑ 2002-2014
<b>Emakakael</b>		↑ 21-57, ↓ 65-85+		↓ 1971-1981, ↑ 2002-2009, ↓ 2012-2016
<b>Eesnääre</b>	↑ 44-77, ↓ 80-85+		↑ 1974-2012, ↓ 2014-2016	

Antud analüüsis on kasutatud splaine, mistõttu loodud mudelite põhjal on võimalik teha ainult lühiajalisi prognoose – umbes 1–5 aastat. Magistritöö autori arvates tuleks vähki haigestumise arvu pikemaajaliseks prognoosimiseks (näiteks 10–20 aastat ette) kasutada aegridu. Selleks on võimalik vähiregistrist taotleda andmeid diagnoosimise kuu ja aasta täpsusega. Küll on aga antud töö metoodika põhjal võimalik väga hästi hinnata vähki haigestumise riski vanuse ja soo järgi.

Edaspidi on Tervise Arengu Instituudil plaanis teha sarnaseid analüüse ja välja arendada erinevaid meetodikaid, et uurida ja prognoosida vähki haigestumist. Tänu nendele analüüsidele on võimalik riskirühmad kindlaks teha ja missugustele vähiliikidele tuleks tulevikus rohkem tähelepanu pöörata.



## Sissejuhatus

Vähihaigestumust Eestis on seni peamiselt uuritud kirjeldava analüüsi abil. Tervise Arengu Instituut on regulaarselt avaldanud ülevaateid ja artikleid vähi esinemise kohta Eestis. Ivar-Endrik Eiche on oma magistritöös „Eesnäärmevähi haigestumus- ja suremustrendid Eestis“ 2016. aastal uurinud eesnäärmevähi haigestumist aastatel 1995–2013 diagnoositud vähijuhtumite alusel. Eiche vaatles oma magistritöös ka statistiliselt olulisi haigestumuskordaja muutusi, kuid vaid grupeeritud andmete alusel. [6] Detailseid vähi haigestumise riski muutumise uuringuid, kuhu oleks kaasatud kõik alates 1968. aastast vähiregistrist kättesaadavad andmed, pole Eestis veel tehtud.

Antud uuringus on autor soovinud luua prognoosimudelid kaasates kõiki analüüsi valmimise ajal kättesaadavaid vähiregistri andmeid aastatel 1968–2017. Lisaks on haigestumise riski leidmiseks kasutatud rahvaaru andmeid Statistikaameti andmebaasist. Lähemalt on uuritud vähi haigestumist kõikide paikmete korral ning soo järgi on eraldi võrreldud naiste puhul emakakaelavähki ja meeste puhul eesnäärmevähki. Selline valik on peamiselt tehtud seetõttu, et vaadelda võimalikult erineval kujul olevaid andmeid, mis võimaldab peale uuringu valmimist luua mudeleid sarnaselt ka kõikidele teistele vähipaikmetele.

Antud analüüsis on kasutatud vähi haigestumise riskimudeli loomisel Poissoni või kvaasi-Poissoni regressiooni. Mudelisse on muutujatena kaasatud patsiendi vanus, diagnoosimise aasta ja võimalusel patsiendi sugu. Lisaks on kasutatud patsiendi soo ja vanuse ning soo ja diagnoosimise aasta koosmõju. Patsiendi vanuse ja diagnoosimise aasta puhul on analüüsis kasutatud splaine mudelite paindlikkuse tagamiseks. Modelleeritavate andmete Clopper–Pearson'i usaldusintervallide leidmiseks on kasutatud *binom.test* funktsiooni R-is. Seost kirjeldava funktsiooni kasvamis-, kahanemis- ja käänupunktide tuvastamiseks on kasutatud  $\chi^2$ -testi või *t*-testi. Lisaks on standardiseeritud 1968.–2015. aasta vähijuhtumite arvud 2016. aasta rahvastiku soo ja vanusjaotuse järgi.

Antud uuringu eesmärkideks on:

1. teada saada, kuidas vanus ja sugu mõjutavad vähi haigestumise riski ja kirjeldada, kuidas haigestumus on ajas muutunud;
2. teada saada, kas ja kui hästi saab antud meetodika põhjal loodud mudelit kasutada tulevaste haigestumisjuhtude prognoosimiseks;
3. teada saada, kas vähi haigestumiste arv kõikide paikmete korral on kasvanud või kahanenud, kui rahvastiku sooline ja vanuseline jaotus oleks kõikidel aastatel sama;
4. teada saada, missuguste aastate ja vanuste korral esineb andmetes seost kirjeldava funktsiooni põhjal statistiliselt olulised kasvamis- ja kahanemised.

Töö jaguneb neljaks peatükiks. Esimeses peatükis tutvustatakse lähemalt analüüsis kasutatud andmeid ja nende ülesehitust. Lisaks on lühidalt kirjeldatud, mis haigus on vähk ja toodud välja analüüsis käsitletud vähiliigid. Teises peatükis kirjeldatakse magistritöös kasutatud meetodikaid. Töö kolmas ja neljas peatükk on analüüsi praktiline osa, kus antakse modelleeritavate andmete üldine ülevaade ja kirjeldatakse loodud mudeleid. Kolmandas peatükis on välja toodud, kuidas summaarselt vähi haigestumise risk muutub inimese vanuse ja haiguse diagnoosimise aasta järgi. Lisaks on uuritud vähijuhtude arvu vanuse järgi 2000. aastal ja diagnoosimise aasta järgi 60aastaste isikute hulgas. Neljandas peatükis on kirjeldatud loodud mudeleid ja tehtud läbi eelnevalt kirjeldatud analüüs vähi haigestumise riski uurimiseks erinevate paikmete korral. Et uurida joonise abil inimese vanuse mõju riskile vähi haigestuda, on fikseeritud aasta 2000 ja diagnoosimise aasta mõju uurimiseks on fikseeritud patsiendi vanus 60.

Töös on andmete töötlemiseks, sh ühendamiseks ning jooniste, kirjeldava analüüsi ja mudelite loomiseks kasutatud rakendustarkvara R. Töös on läbivalt olulisusnivooks võetud  $\alpha = 0.05$  ja vastavaks usaldusnivooks  $1 - \alpha = 0.95$ .

# 1 Andmed

## 1.1 Vähiregister

Eesti vähiregistri põhiülesanne on tagada võimalikult täielik ja usaldusväärne vähijuhtude registreerimine [32].

Vähiregister asutati 1978. aastal, ent registris on olemas andmed kõigi Eestis alates 1968. aastast diagnoositud vähijuhtude kohta. Alates 1994. aastast kuuluvad registreerimisele lisaks *in situ* kasvaja<sup>1</sup> ja 1998. aastast peaaegu ja kesknärvisüsteemi ning peaaegu piirkonnas paiknevate sisesekretsiooninäärmete healoomulised ja ebaselge või teadmata loomusega kasvaja. [32]

Vähijuhtude andmete täielikkuse ja kvaliteedi tagamisel on ülioluline andmete saamine paljudest eri allikatest. Vähijuhtudest teatamine on kohustuslik kõigile Eestis töötavatele arstidele, kes diagnoosivad või ravivad vähki. Samuti esitavad registrile andmeid kohtuarst-ekspertid. [32]

Registris toimub andmevahetus Eesti rahvastikuregistriga (isikuandmed, sh isiku riigist lahkumise aeg) ja surma põhjuste registriga (surmaandmed, sh surmapõhjus). Lisaks kõrvalt vähiregister oma andmebaasi regulaarselt kahe suurhaigla, Põhja-Eesti Regionaalhaigla ja Tartu Ülikooli Kliinikumi, andmebaasidega. [32]

Vähiregister kuulub alates 1989. aastast Rahvusvahelisse Vähiregistrite Assotsiatsiooni (IACR) ning Euroopa Vähiregistrite Võrgustikku (ENCR) [32].

### 1.1.1 Kasutatud andmed

Vähiregister haldab järgmiseid andmeid:

- patsiendi nimi, isikukood, sugu, sünniaeg, sünnikoht, elukoht diagnoosimise hetkel ja rahvus;
- diagnoos ehk üksikasjalik paige, diagnoosimise aeg, diagnoosi kinnitanud uurimismeetodid, morfoloogiline diagnoos ja pahaloomulisuse aste, kasvaja levik;
- esmane ravi raviliigiti – ravi olemus, ravimise koht, ravi alustamise kuupäev ja ravimeetod;
- surmaaeg ja -põhjus või Eestist lahkumise aeg;
- vähiteatise vormistaja andmed. [32]

Käesolevas magistritöös on vaatluse alla võetud:

- patsiendi sugu („Mees“, „Naine“);
- vanus diagnoosimise hetkel („0“, „1“, ..., „84“, „85+“);
- diagnoosi kood RHK-10 klassifikatsiooni alusel (nt „C07“);
- diagnoosimise aasta (1968–2017).

Käesoleva magistritöö valmimise ajal olid vähiregistris kättesaadavad andmed kuni aastani 2017. Samuti polnud võimalik analüüsi kaasata patsiendi elukoha maakonda, kuna Eesti haldusjaotus on olnud pidevas muutumises aastatel 1968–2017 ning seda töömahukat ühildust ei ole teostatud.

Töö autor võttis vastu eksperthinnanguna otsuse ühendada kõik 85aastased ja vanemad vähki haigestunud ühte vanuserühma. See eksperthinnang põhineb kogemusel, kus koostatud mudelite

---

<sup>1</sup> Normist kõrvalekaldunud rakkude rühm, mis pole esialgselt tekkekohast edasi levinud. Need rakud võivad muutuda vähiks ja levida lähedal asuvasse normaalsesse kudedesse. Nimetatakse ka 0-stadiumi vähiks. [22]

interpreteerimisel tekib probleem just 85aastaste ja vanemate isikute puhul (nt 100%-lise tõenäosusega vähki haigestuvad inimesed mingis konkreetses vanuses).

Vähiregistris on andmed järgmiste kasvajate kohta rahvusvahelise haiguste klassifikatsiooni 10. versiooni ehk RHK-10<sup>2</sup> alusel:

- kõik pahaloomulised kasvajad (RHK-10 koodid C00–C97);
- *in situ* kasvajad (D00–D09);
- peaaaju ja kesknärvisüsteemi, samuti peaaaju piirkonnas paiknevate sisesekreetsiooninäärmete healoomulised ning ebaselge ja määratlemata loomusega kasvajad (D32.0–D33.9, D35.2–D35.4, D42.0–D43.9, D44.3–D44.5);
- lümfoid- ja vereloomekoe ning nendele sarnaste kudede muud kasvajad (D45–D47). [32]

Käesolevas magistritöös on vaatluse alla võetud ainult pahaloomulised kasvajad (C00–C97).

Vähiregistri andmete analüüsiks ettevalmistamise R koodi on näha lisas 2.

## 1.1.2 Analüüsis käsitletud vähiliigid

Vaadeldud on üleüldist vähki haigestumist – nende andmete korral on haigestumiste arv ja risk aastatega kasvanud (vt jooniseid 1 ja 4). Lisaks võeti detailsema vaatluse alla emakakaela- ja eesnäärmevähki haigestumine. Nende vähkide esinemist iseloomustanud järgmised trendid:

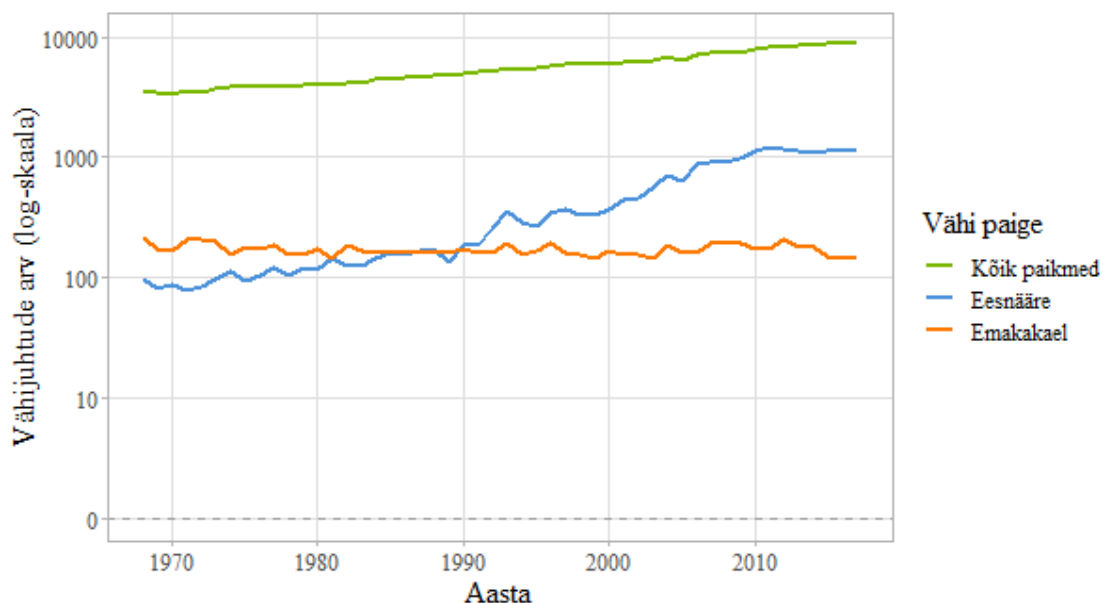
- eesnäärmevähi puhul on lühikese aja jooksul tekkinud järsk haigestumisriski tõus (vt jooniseid 1 ja 10);
- emakakaelavähi juhtude arv ja haigestumisrisk on olnud juhuslikus muutumises ning haigusjuhtude arv on olnud suhteliselt väike (vt jooniseid 1 ja 7).

Nii on uuringu läbiviijal võimalik koostada mudelid, võttes aluseks erineva suundumusega haigestumusandmed, ja rakendada sama meetodikat ka kõikidele teistele vähipaikmetele. See on oluline ainuüksi seetõttu, et kui mõne vähipaikme korral esineb juhtumeid kas ainult meestel või ainult naistel, nagu on valitud eesnäärme- ja emakakaelavähi puhul, siis mudelis pole vajalik rakendada soo tunnust. Kõikide paikmete korral on rakendatud mudelis ka sootunnust, mistõttu seda meetodikat on võimalik rakendada ka ülejäänud vähiliikidele, mis esinevad mõlema soo korral.

Joonistel 1 ja 2 on vähijuhtude arvu uuritud logaritmskaala abil. Logaritmi võtmine ei muuda andmete üldist käitumist, sest arvude kahanemised ja kasvumised jäävad samaks. Haigusjuhtude arv ei ole nendel joonistel standardiseeritud rahvastiku soo- ja vanusjaotuse järgi.

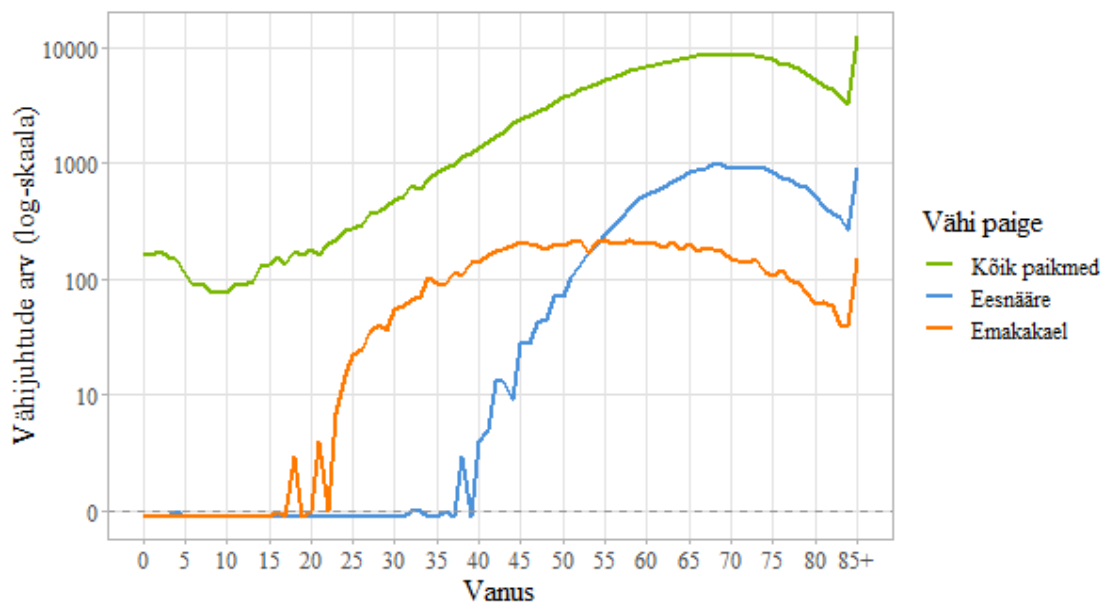
---

<sup>2</sup> Haiguste klassifikatsioon ehk jaotiste süsteem, milles haigused (või haiguste nimetused) on määratud vastavalt kehtestatud kriteeriumitele. RHK-10 peamine eesmärk on võimaldada eri aegadel kogutud suremuse ja haigestumuse andmete süstemaatiline registreerimine, analüüsimine, interpreteerimine ja võrdlemine. Seda klassifikatsiooni kasutatakse haiguste diagnooside ja muude terviseprobleemide ülekandmiseks sõnadest täkkoodi. See võimaldab andmete hõlpsat säilitamist, otsingut ja analüüsi ka rahvusvahelisel tasandil. [1]



**Joonis 1.** Vähijuhtude arvu muutumine sõltuvalt aastast valitud paikme järgi, 1968–2017

Jooniselt 1 on näha, et vähijuhtude arv on kasvanud, kui vaadata kõiki paikmeid kokku ja samuti eesnäärme puhul. Emakakaelavähi puhul on juhtude arv aastate jooksul olnud pidevas muutumises, kuid jäänud enam-vähem samasse suurusjärku.



**Joonis 2.** Vähijuhtude arvu muutumine sõltuvalt vanusest valitud paikme järgi, 1968–2017

Jooniselt 2 on näha, et juhtumite arv on noores eas väike nii eesnäärme- kui ka emakakaelavähi puhul. Nende paikmete korral kasvab vähijuhtude arv teatud vanusest alates oluliselt. Juhtude arv kasvab märkimisväärselt eesnäärmevähi puhul vanusevahemikus 35–45 aastat ja emakakaelavähi puhul 15–25 aastat. Vanuse kasvades see suurenemine aeglustub ning kõrges vanuses juhtude arv taas väheneb.

Kõiki paikmeid kokku vaadeldes on näha, et varases eas esineb mõnevõrra rohkem vähijuhte kui vanuses 5–10 eluaastat, kuid uuritava tunnuse väärtused hakkavad kasvama alates 10. eluaastast. Alates 70. eluaastast on vähijuhtude arvus näha uuesti kerget vähenemist.

Vanuses „85+“ juhtude arv on suur, sest sellesse kuuluvad kõik haigestumise juhud, mis esinesid 85aastastel ja vanematel inimestel ja see moodustab suurema grupi.

#### *1.1.2.1 Mis on vähk?*

Vähktõbi on haigus, mida iseloomustab keharakkude peatumatu paljunemine ja levimine ümbritsevasse kudedesse [25]. Vähiraku eelis teiste rakkude ees on selle pidurdamatu kasv ja pikem eluiga [15]. Inimkeha koosneb triljonitest rakkudest ja vähk võib tekkida kõikjal. Inimesel tekivad uued rakud peamiselt siis, kui keha neid vajab. Kui rakud vananevad või saavad kahjustada, siis nad surevad ja nende asemel tekivad uued. Vähk tekib siis, kui see korrapärane protsess läheb paigast ära. Vähi tekkides ei hävitata kahjustunud või vanu rakke ja juurde tekivad lisaks uued rakud, kuigi neid pole vaja. Need uued lisarakud võivad hakata peatumatult paljunema ja seetõttu tekivadki kasvaja. [25]

Kasvaja võivad olla healoomulised või pahaloolumulised. Pahaloolumuliste kasvaja rakkudel on omadus kasvada läbi normaalse kude ja seeläbi organismis levida. Lisaks võivad nende rakud liikuda vere või lümfisüsteemi kaudu keha kaugematesse kohtadesse ja moodustada algsest kasvajast kaugel uued kasvaja. [25] Lisaks mõjutavad vähirakud mikrokeskkonda ja immuunsussüsteemi põhikomponente, mille tõttu suudavad nad vältida hävinemist, ellu jääda näiliselt ebasoodsas keskkonnas ning kahjustada ka teisi elundeid [15]. Erinevalt pahaloolumulistest kasvajatest ei levi healoomulised kasvaja lähedalasuvatesse kudedesse. Healoomulised kasvaja võivad mõnikord siiski olla üsna suured. Kasvajakollete eemaldamisel kirurgiliselt healoomulised kasvaja tavaliselt tagasi ei kasva, kuid pahaloolumulised kasvaja võivad seda mõnikord teha. Üldjuhul ei ole healoomulised kasvaja eluohtlikud (v.a ajukasvaja). [25]

Enamik pahaloolumulistest kasvajatest on nimetuse saanud organi järgi, kus nad tekivad (nt rinnavähk, kopsuvähk jms). Kasvaja iseloomustatakse ka nende ehituse ehk histoloogia järgi (nt lamerakk kartsinoom). Lisaks sellele iseloomustatakse kasvaja stadiumitega nende leviku ulatuse järgi. [30]

Kokku on 1968.–2017. aastal registreeritud 279 416 erinevat vähijuhtumit. Pahaloolumuliste kasvaja grupe ja vähijuhtude esinemissagedusi on võimalik näha lisas 1.

#### *1.1.2.2 Emakakaelavähk (C53)*

Emakakael on emaka, õõnsa, pirnukujulise elundi, kus rasestumise korral kasvab loode, alumine ja kitsas osa, mis ühendab emakat tupega. Emakakaelavähk on haigus, mille korral emakakaelas moodustuvad pahaloolumulised vähirakud. Haigus esineb ainult naistel. [21]

Emakakaelavähk tekib tavaliselt aeglaselt. Enne vähi ilmnemist emakakaelas läbivad emakakaela rakud rea muutusi, mille käigus emakakaela kudedesse hakkavad tekkima normist kõrvalekalduvad rakud – seda protsessi nimetatakse düsplaasiaks. Emakakaela ebatüüpilised rakud võivad ilma ravita kaduda, jääda samaks või muutuda mitmete aastate jooksul vähirakkudeks. [21] Emakakaelavähk levib esmalt ümbritsevatele kudedele, lümfiteede kaudu vaagna lümfisõlmedele ja kõhuaorti ümbritsevatele lümfisõlmedele. [14] Tavalisteks sümptomiteks on metrorrhagia<sup>3</sup>, sünnitusjärgne määrimine, emakakaela haavandid ning kusepõie ja pärasoole talitlushäired, millele järgnevad fistulid ja valud [26].

---

<sup>3</sup> Suurema hulga verega menstruaalverejooksud [23].

### **Teadaolevad riskifaktorid:**

- papilloomiviirus (HPV);
- suitsetamine;
- HIV ja muudest teguritest põhjustatud vähenenud immuunsus;
- varajane seksuaaleluga alustamine ja rohke seksuaalpartnerite arv;
- toitainete puudus (A-vitamiini puudus). [21][26]

Emakakaelavähki võib vaadelda kui suguliselt levivat haigust. Kui naine pole kunagi olnud seksuaalvahekorras, on peaaegu olematu võimalus, et ta sellesse vähki haigestub. See on peamiselt seotud asjaoluga, et vaid seksuaalvahekorra ajal on võimalik nakatuda inimese papilloomiviirusega (HPV). Uuringud on tõestanud, et HPV-ga nakatumise risk on suurem naistel, kes on olnud seksuaalselt aktiivsed enne 18. eluaastat ja kellel on olnud enam kui 6 seksuaalpartnerit. Kuid tuleb mainida, et enam kui 80-st HPV tüübist ainult ligi 30 võivad põhjustada emakakaelavähki. [21]

#### *1.1.2.3 Eesnäärmevähk (C61)*

Eesnääre on meeste reproduktiivsüsteemi nääre. Eesnääre asub vahetult põie all ning pärasoole ees – see on uriini koguv ja tühjendav organ. See on umbes kreeka pähkli suurune ja ümbritseb kusiti osa. Eesnääre toodab vedelikku, mis moodustab osa spermast. Eesnäärmevähk on haigus, mille korral pahaloomulised vähirakud moodustuvad eesnäärme kudedes. Haigus esineb ainult meestel. [24]

Enamik eesnäärmevähi juhtumeid saavad alguse näärmelisest koest. Eesnäärmevähk kasvab haiguse algfaasis aastate jooksul aeglaselt. [5] Mitmete faktorite, sealhulgas suguhormoonide, koosmõjul toimuvad eesnäärme rakkude DNAs muutused, mis põhjustavad rakkude pidurdamatut paljunemist [24]. Varem või hiljem kasvab eesnäärmevähk siiski läbi eesnäärme kapsli ümbritsevatesse kudedesse [5].

Eesnäärmevähk, nagu kõik pahaloomulised kasvaja, võib levida organismis ja anda siirdeid ehk metastaase teistesse organitesse. [24] Sageks metastaaside kohaks võib olla luu, mille puhul võivad esineda valud või patoloogilised murrud. Tavaliselt tekivad sümptomid peamiselt alumistes kuseteedes: näiteks sage, kiireloomuline või kaheldav urineerimisevajadus, öine põiepidamatus ning mõnikord veri uriinis. Samuti võib esineda erektsioonihäired ja verd spermas. [2]

### **Olulisemad riskifaktorid:**

- kõrge vanus;
- geneetiline eelsoodumus;
- aafrika päritolu;
- ülekaal. [13][17]

Suurel osal meestest areneb elu jooksul eesnäärmevähk, kuid mehed sageli surevad mingil muul põhjusel. Seega tugevaim riskifaktor näib olevat vanus – mida kauem mees elab, seda suurema tõenäosusega diagnoositakse tal elu jooksul eesnäärmevähk. Meestel, kellel on esimese astme sugulastel (isa, poeg, vend) diagnoositud eesnäärmevähk, on suurem tõenäosus haigestuda elu jooksul samasse haigusse. [24]

## 1.2 Statistikaameti andmebaas

Eesti Statistikaameti andmebaas esitab riiklikku statistikat. Riiklik statistika hõlmab andmeid, mida Statistikaamet avaldab Vabariigi Valitsuse korraldusega kinnitatud riikliku statistika programmi alusel. Statistikaameti andmebaasis esitatakse kogu avaldatav statistika. [4]

Statistikaameti andmebaas esitab statistikat nelja statistika põhivaldkonna – keskkond, majandus, rahvastik ja sotsiaalelu – kohta. Rahva ja eluruumide loenduse ning põllumajandusloenduse andmed on esitatud eraldi. Valdkonnad jagunevad alamvaldkondadeks, moodustades puustruktuuri. Valdkonnad ja alamvaldkonnad on esitatud tähestiku järjekorras. [4]

Andmed on andmebaasis esitatud tabelitena, millel on unikaalne kood. Valdkonnatabelite loetelu on sorditud tabeli koodi järgi. [4]

Käesolevas magistritöös on kasutatud järgnevate tabelite andmeid:

- RV02111: Rahvastik soo ja vanuse järgi, 1. jaanuar (1950–1999)
- RV0212: Rahvastik aasta alguses ja aastakeskmise rahvaarv soo ja vanuse järgi

Esimesest tabelist on võetud andmed 1968.–1999. aasta kohta soo („Mees“, „Naine“) ja vanuse („0“, „1“, ... „84“, „85+“) järgi. Teisest tabelist on võetud sama ülesehitusega andmed, kuid 2000.–2017. aasta kohta. Antud uuringus kasutatakse rahvaarve 1. jaanuari seisuga järgi.

Statistikaameti rahvaarvu ja vähiregistri andmete ühendamise R koodi on näha lisa 2.

## 2 Metoodika

### 2.1 Mudel

#### 2.1.1 Poissoni mudel

Käesolevas magistritöös on kasutatud üldistatud lineaarset **Poissoni mudelit**. See mudel tuleb kasuks, kui luuakse prognoose suurusele  $y_i$ , mis on ülimalt loenduv suurus<sup>4</sup> [28]. Antud analüüsis on loendatud vähijuhtude arvu.

Poissoni tõenäosusjaotus on kujul:

$$f(y) = \frac{\lambda^y e^{-\lambda}}{y!}, \quad y = 0, 1, 2, \dots$$

Poissoni regressioonimudel on kujul:

$$y_i = E(y_i) + \epsilon_i, \quad i = 1, 2, \dots, n,$$

kus  $y_i$ -d on sõltumatud Poissoni jaotusega juhuslikud suurused ja  $E(y_i) = \lambda_i$  on funktsioon kujul:

$$\lambda_i = e^{\mathbf{x}'_i \boldsymbol{\beta}} = e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}, \text{ kus } \mathbf{x}_i = \begin{pmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \end{pmatrix} \text{ ja } \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}. \quad [28]$$

Käesolevas magistritöös on mudeli konstrueerimiseks kasutatud rakendustarkvara R funktsiooni `glm`: `model <- glm(..., family=poisson())` (vt lisa 4).

#### 2.1.2 Grupeeritud andmed

Käesolevas magistritöös on tegemist vanuse, aasta ja soo põhjal grupeeritud haigestumisandmetega. Antud analüüsi peamine huvi on teada saada, kui suur on risk haigestuda vähki. Selleks, et riski arvutada, peab teadma iga rahvastikugrupi korral inimeste koguarvu, kelle haigestumise riski soovitakse hinnata [3]. Tähistagu  $i$  grupi tunnust, mis on määratud patsiendi vanuse ja soo ning haiguse diagnoosimise aasta järgi ja  $n_i$  rahvaarvu grupis  $i$ .

Olgu uued tähistused:

- $y_{ij}$  – inimene  $j$  grupis  $i$ , mille väärtus on 0 (terve) või 1 (haigestunud);
- $y_i = \sum_j y_{ij}$  – kõikide haigestunute arv grupis  $i$ . [19]

Eeldame, et haigestumise tõenäosus on kõigil samasse gruppi kuuluvatel inimestel sama,  $y_{ij} \sim \text{Bin}(1, \lambda_i)$  ( $j = 1, 2, \dots, n_i$ ), millest eeldades juhuslike suuruste sõltumatust järeldub, et  $y_i \sim \text{Bin}(n_i, \lambda_i)$ . Suure valimi korral  $y_i \sim \text{Po}(\lambda_i n_i)$  ja  $E(y_i) = n_i \lambda_i$ . [20]

Mudeli kuju grupeeritud andmete korral on:

$$\begin{aligned} \ln(E(y_i)) &= \ln(n_i \lambda_i) = \ln(n_i) + \ln(\lambda_i) \stackrel{\lambda_i = e^{\mathbf{x}'_i \boldsymbol{\beta}}}{=} \ln(n_i) + \mathbf{x}'_i \boldsymbol{\beta} \\ \Rightarrow E(y_i) &= e^{\ln(n_i) + \mathbf{x}'_i \boldsymbol{\beta}} = e^{\ln(n_i)} \cdot e^{\mathbf{x}'_i \boldsymbol{\beta}} = n_i \cdot e^{\mathbf{x}'_i \boldsymbol{\beta}} \\ \Rightarrow \frac{E(y_i)}{n_i} &= e^{\mathbf{x}'_i \boldsymbol{\beta}}. \quad [19][31] \end{aligned} \quad (1)$$

<sup>4</sup> ülimalt loenduv suurus – lõplik või loenduv suurus



Suurus  $\frac{E(y_i)}{n_i}$  ongi haigestumise riski hinnang. Edaspidi tähistagu  $\mu_i := E(y_i) = n_i \cdot e^{x_i' \beta}$ .

Käesolevas magistritöös on mudeli loomiseks grupeeritud andmete korral kasutatud rakendustarkvara R funktsiooni glm kujul:

`model <- glm(..., offset=log(Rahvaarv))` (vt lisa 4).

## 2.1.3 Mudeli hindamine

Igal juhul peavad suurused  $\mu_i$  olema positiivsed. Et hinnata mudeli parameetreid  $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$ , võib kasutada suurima tõepära hinnangut. Tõepärafunktsioon on grupeeritud andmete korral kujul:

$$L(\beta) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}. \quad [19]$$

Antud tõepärafunktsiooni log-tõepära on:

$$\begin{aligned} l(\beta) &= \ln(L(\beta)) = \ln\left(\prod_{i=1}^n f(y_i)\right) & (2) \\ &= \ln\left(\prod_{i=1}^n \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}\right) = \sum_{i=1}^n \ln\left(\frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}\right) \\ &= \sum_{i=1}^n (y_i \ln(\mu_i) - \mu_i - \ln(y_i!)) = \sum_{i=1}^n (y_i \ln(\mu_i) - \mu_i - \ln(y_i!)) \\ &\stackrel{\mu_i = n_i \cdot e^{x_i' \beta}}{\cong} \sum_{i=1}^n (y_i \ln(n_i \cdot e^{x_i' \beta}) - n_i \cdot e^{x_i' \beta} - \ln(y_i!)) \\ &= \sum_{i=1}^n (y_i \ln(n_i) + y_i \ln(e^{x_i' \beta}) - n_i \cdot e^{x_i' \beta} - \ln(y_i!)) \\ &= \sum_{i=1}^n y_i \ln(e^{x_i' \beta}) - \sum_{i=1}^n n_i e^{x_i' \beta} + \underbrace{\sum_{i=1}^n (y_i \ln(n_i) - \ln(y_i!))}_{Const} \\ &= \sum_{i=1}^n y_i x_i' \beta - \sum_{i=1}^n n_i e^{x_i' \beta} + Const. \quad [9][19] \end{aligned}$$

Skoorifunktsioon on kujul:

$$\begin{aligned} s(\beta) &= \frac{\partial l(\beta)}{\partial(\beta)} = \frac{\partial (\sum_{i=1}^n y_i x_i' \beta - \sum_{i=1}^n n_i e^{x_i' \beta})}{\partial(\beta)} \\ &= \sum_{i=1}^n y_i x_i - \sum_{i=1}^n n_i e^{x_i' \beta} \cdot x_i = \sum_{i=1}^n (y_i - n_i e^{x_i' \beta}) x_i \end{aligned}$$

$$\mu_i = n_i e^{x_i' \beta} \quad \hat{=} \quad \sum_{i=1}^n (y_i - \mu_i) x_i. \quad [9][19] \quad (3)$$

Võrdsustades skoorifunktsiooni 0-ga saab leida suurima tõepära hinnangu  $\hat{\beta}$ :

$$s(\beta) = 0 \quad \Rightarrow \quad \sum_{i=1}^n (y_i - n_i e^{x_i' \hat{\beta}}) x_i = 0. \quad [9][19] \quad (4)$$

Informatsioonimaatriks on kujul:

$$I(\beta) = -E \left( \frac{\partial^2 l(\beta)}{\partial(\beta) \partial(\beta')} \right) = \sum_{i=1}^n n_i e^{x_i' \hat{\beta}} x_i x_i' \quad \stackrel{\mu_i = n_i e^{x_i' \hat{\beta}}}{\hat{=}} \quad \sum_{i=1}^n \mu_i x_i x_i'.$$

Informatsioonimaatriksist saab asümptootilise kovariatsioonimaatriksi suurima tõepära hinnangule  $\hat{\beta}$ :

$$D(\hat{\beta}) = \left( \sum_{i=1}^n \hat{\mu}_i x_i x_i' \right)^{-1}. \quad [9]$$

## 2.1.4 Ala- ja ülehajuvus

Olgugi, et Poissoni mudel on lihtsasti interpreteeritav mudel, siis selle eeldused võivad olla veidi liiga kitsendavad. Täpsemalt on selle mudeli eelduseks see, et prognoositava ehk uuritava tunnuse keskväärtus on võrdne selle dispersiooniga:

$$D(y_i) = E(y_i) = \mu_i. \quad [9] \quad (5)$$

Kui uuritava tunnuse dispersioon on palju suurem kui selle keskväärtus, siis öeldakse, et andmed on ülehajuvuses. Harvematel juhtudel esineb alahajuvust ehk dispersioon on palju väiksem kui keskväärtus. [9]

Ala- ja ülehajuvuse uurimiseks võetakse kasutusele parameeter  $\phi$  ja kirjutatakse:

$$D(y_i) = \phi \mu_i. \quad (6)$$

Parameetri  $\phi > 0$  abil saab hinnata potentsiaalset ala- ja ülehajuvust:

- Kui  $\phi = 1$ , siis ala- või ülehajuvust ei esine;
- Kui  $\phi > 1$  ehk  $D(y_i) > \mu_i$ , siis esineb ülehajuvus;
- Kui  $\phi < 1$  ehk  $D(y_i) < \mu_i$ , siis esineb alahajuvus. [9][19]

Valemist (3) on võimalik näha, et mudeli parameetrite  $\beta$  korrektseks hindamiseks piisab, kui uuritava tunnuse keskmise funktsioon  $\mu_i = E(y_i)$  on hinnatud õigesti. Seetõttu mudeli hinnangule  $\hat{\beta}$  viidatakse tihti ka kui kvaasitõepära hinnangfunktsioonile. Selle hinnangfunktsiooniga on võimalik leida hinnangulised keskmised  $\hat{\mu}_i$  ja seejärel hinnata parameetrit  $\phi$  järgmiselt:

$$\hat{\phi} = \frac{1}{n - (k + 1)} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}. \quad [9]$$

Ülehajuvuse korral ja eeldades võrrandi (6) kehtivust rakendatakse kvaasi-Poissoni regressiooni, mille korral skaleeritakse ümber mudeli parameetrite hinnangute standardhälbed kasutades hinnatud ülehajuvusparameetrit. Kvaasi-Poissoni jaotust otseselt ei eksisteeri, kuid kvaasi-Poissoni regressioon võimaldab leida korrektsemaid dispersioonihinnanguid mudeli parameetrite hinnangutele. [9]

Käesolevas magistritöös on üle- ja alahajuvuse leidmiseks kasutatud reeglit:

$$\text{suhe} = \frac{\chi^2}{df} \approx 1. \quad (7)$$

Kui suhe  $> 2$ , siis on kasutatud kvaasi-Poissoni regressiooni. [19]

Käesolevas magistritöös on kasutatud ülehajuvuse uurimiseks ja kvaasitõepäraga mudeli konstrueerimiseks rakendustarkvara R funktsiooni:

```
sum(residuals(mudel, type="pearson")^2)/mudel$df.residual
mudel <- update(mudel, family=quasipoisson("log")) (vt lisa 5).
```

## 2.2 Splainid

Kasutades polünoomset regressiooni on mudeleid küll lihtne interpreteerida, kuid nende rakendamisel võivad esineda teatud raskused. Näiteks kui argumenttunnuse muutumispiirkonna teatud osas on lokaalne kõrvalekalle üldisest trendist, siis taoliste kõrvalekallete modelleerimiseks tuleb kasutada väga kõrge astmega polünoomi. Splainid pakuvadki kompromissi esitades mudelit piirkonniti erinevate polünoomidena, mis muudavad mudelid palju paindlikumaks. Muutuja piirkonnad on määratud murdepunktidega  $\xi_1, \dots, \xi_K$ . Lisaks on oluline, et polünoomid liituks omavahel sujuvalt murdepunktides. Kõige enam kasutatakse splainide puhul kuuppöloome, kus murdepunktides on pidev esimene ja teine tuletis. Samasugune valik on tehtud ka antud analüüsis. Inimese silm on võimeline nägema murdepunkte esimese ja teise astme, kuid mitte kõrgema astme polünoomide korral. [11]

Peamine raskus splainide puhul on mudeli murdepunktide asukohta väljaselgitamine. Murdepunktide genereerimise viise on lõpmata palju. Näiteks murdepunktide asukohta võib genereerida kasutades ühtlast jaotust, mille parameetriteks on muutuja väärtuste vahemiku minimaalne ja maksimaalne väärtus. Käesolevas analüüsis on murdepunktide asukoht leitud aga kvantiilide järgi – näiteks kui murdepunktide arv on 3, siis on kasutatud mudeli murdepunktideks muutuja 0,25-kvantiili, 0,5-kvantiili ja 0,75-kvantiili. [11]

Olgu antud muutuja murdepunktid  $\xi_1, \dots, \xi_K$  nii, et  $\xi_1 < \dots < \xi_K$ . Üldiselt on mudeli kujuks:

$$s(x) = \beta_0 + \dots + \beta_m x^m + \sum_{i=1}^K \theta_i (x - \xi_i)_+^m,$$

kus  $a_+ = \begin{cases} a, & \text{kui } a \geq 0 \\ 0, & \text{kui } a < 0 \end{cases}$ . [11][31] Seda suurust võib kirjutada ka kujul  $I(a \geq 0)$ :

$$s(x) = \beta_0 + \dots + \beta_m x^m + \sum_{i=1}^K \theta_i (x - \xi_i)^m \cdot I(x - \xi_i \geq 0).$$

Käesolevas magistritöös on uuritava tunnuse mudeli konstrueerimisel kasutatud tunnuseid Vanus, Aasta ja Sugu ning koosmõjusid Vanus\*Sugu ja Aasta\*Sugu.

Olgu  $n$  tunnuse vanus ja  $m$  tunnuse Aasta murdepunktide arv ning olgu  $\xi_1, \dots, \xi_n$  tunnuse Vanus ja  $\eta_1, \dots, \eta_m$  tunnuse Aasta murdepunktid. Tähistagu:

$$\begin{aligned} x &:= \text{Vanus} \\ y &:= \text{Aasta} \\ z &:= \text{Sugu} \end{aligned} \quad (8)$$

Antud analüüsis  $z = 1$ , kui tegemist on naisega ja  $z = 0$ , kui tegemist on mehega.

Antud analüüsis käsitletud mudeli lineaarseks osaks on:

$$\begin{aligned}
 s(x, y, z) = & \beta_0 + \beta_1 z + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 \\
 & + \sum_{i=1}^n \beta_{4+i} (x - \xi_i)^3 \cdot I(x - \xi_i \geq 0) \\
 & + \beta_{5+n} y + \beta_{6+n} y^2 + \beta_{7+n} y^3 + \sum_{i=1}^m \beta_{7+n+i} (y - \eta_i)^3 \cdot I(y - \eta_i \geq 0) \\
 & + \beta_{8+n+m} (x \times z) + \beta_{9+n+m} (x^2 \times z) + \beta_{10+n+m} (x^3 \times z) \\
 & + \sum_{i=1}^n \beta_{10+n+m+i} ((x - \xi_i)^3 \times z) \cdot I(x - \xi_i \geq 0) \\
 & + \beta_{11+2n+m} (y \times z) + \beta_{12+2n+m} (y^2 \times z) + \beta_{13+2n+m} (y^3 \times z) \\
 & + \sum_{i=1}^m \beta_{13+2n+m+i} ((y - \eta_i)^3 \times z) \cdot I(y - \eta_i \geq 0).
 \end{aligned} \tag{9}$$

Käesolevas magistritöös on kasutatud Poissoni regressiooni mudelit, mis valemi (1) põhjal omandab kuju:

$$\frac{E(y_i)}{n_i} = e^{s(x,y,z)}, \tag{10}$$

kus  $s(x, y, z)$  kuju on välja toodud valemis (9).

Mudeli konstrueerimiseks on vaja hinnata parameetrite vektorit  $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{13+2n+2m} \end{pmatrix}$ .

Splainide koostamiseks on kasutatud rakendustarkvara R funktsiooni bs:

bs(..., df=n, degree=3) (vt lisa 5).

## 2.3 Parima mudeli valimine

Parima mudeli valimiseks on kasutatud Akaike informatsioonikriteeriumit (AIC), mis on arvutatav valemiga:

$$AIC = -2 \cdot l(\hat{\boldsymbol{\beta}}) + 2 \cdot p,$$

kus

- $\hat{\boldsymbol{\beta}}$  on suurima tõepära hinnang mudeli parameetritele  $\boldsymbol{\beta}$ , mis on leitav võrrandist (4);
- $l$  on log-tõepärafunktsioon kujul (2);
- $p$  on parameetrite arv mudelis. [18]

Suurust AIC kasutatakse eelkõige seetõttu, et välja selgitada parameetrite arvu  $p$ . Selleks võrreldakse omavahel mitut alternatiivset mudelit erinevate parameetrite arvuga  $p$  ja valitakse välja see, mis minimiseerib suurust AIC. [9]

AIC abil on uurija püüdnud välja selgitada murdepunktide arvu splainides. Selleks on loodud tsükkel, mis kontrollib, kas murdepunktide lisamine mudelisse vähendab suurust AIC või mitte. Tsükklis on vaadeldud kaks sammu ette – kui ühe murdepunkti lisamine mudelisse ei paranda mudeli AIC väärtust, siis on uuritud lisaks AIC väärtust ka kahe murdepunkti lisamisel. Alguses on eelnevalt kirjeldatud tsükkel rakendatud tunnusele Vanus ja siis tunnusele Aasta.

Parima mudeli valiku R koodi on võimalik näha lisas 4.

## 2.4 Seost kirjeldava funktsiooni kasvamis-, kahanemis- ja käänukohad

Seost kirjeldava funktsiooni kasvamis-, kahanemis- ja käänukohtade identifitseerimiseks on oluline leida mudeli tuletis. Kui tegemist on mudelis käänukohaga ehk selle ekstreemumiga mingis punktis, siis peab mudeli tuletise väärtus selles punktis olema 0. [10] Käänukohtade leidmiseks kontrollitakse hüpoteese:

$$\begin{aligned} H_0: \left( \frac{E(y_i)}{n_i} \right)' &= 0 \\ H_1: \left( \frac{E(y_i)}{n_i} \right)' &\neq 0 \end{aligned} \quad (11)$$

Lisaks kehtivad:

$$\begin{aligned} \left( \frac{E(y_i)}{n_i} \right)' > 0 &\Rightarrow \text{vaadeldavas piirkonnas seost kirjeldav funktsioon kasvab;} \\ \left( \frac{E(y_i)}{n_i} \right)' < 0 &\Rightarrow \text{vaadeldavas piirkonnas seost kirjeldav funktsioon kahaneb.} \end{aligned}$$

Mudeli (10) tuletist võib siin vastavalt uurimisküsimusele võtta kas tunnuse vanus või Aasta järgi. Tänu sellele saab näiteks näidata, et kui mudeli tuletis Aasta järgi on positiivne mingil fikseeritud aastal, siis sellel aastal vähki haigestumise risk kasvas.

Mudeli tuletis tunnuse vanus järgi (kõik tähistused on selgitatud osas (8) ja  $s(x, y, z)$  kuju on välja toodud real (9)):

$$\begin{aligned} \frac{d(e^{s(x,y,z)})}{dx} &= \frac{d(e^{s(x,y,z)})}{d(s(x,y,z))} \cdot \frac{d(s(x,y,z))}{dx} \\ &= e^{s(x,y,z)} \cdot \frac{d(s(x,y,z))}{dx} \\ &= \frac{e^{s(x,y,z)}}{>0} \cdot \left( \beta_2 + 2\beta_3x + 3\beta_4x^2 + \sum_{i=1}^n 3\beta_{4+i}(x - \xi_i)^2 \cdot I(x - \xi_i \geq 0) \right. \\ &\quad \left. + \beta_{8+n+m}z + 2\beta_{9+n+m}(x \times z) + 3\beta_{10+n+m}(x^2 \times z) \right. \\ &\quad \left. + \sum_{i=1}^n 3\beta_{10+n+m+i}((x - \xi_i)^2 \times z) \cdot I(x - \xi_i \geq 0) \right). \end{aligned}$$

Kuna  $e^{s(x,y,z)} > 0$ , siis hüpoteeside (11) kontrollimiseks on oluline teada, kas teine tegur on võrdne nulliga.

Sama mudeli tuletis tunnuse Aasta järgi:

$$\begin{aligned} \frac{d(s(x,y,z))}{dy} &= \frac{d(e^{s(x,y,z)})}{d(s(x,y,z))} \cdot \frac{d(s(x,y,z))}{dy} \\ &= e^{s(x,y,z)} \cdot \frac{d(s(x,y,z))}{dy} \end{aligned}$$

$$\begin{aligned}
&= \frac{e^{s(x,y,z)}}{>0} \cdot \left( \beta_{5+n} + 2\beta_{6+n}y + 3\beta_{7+n}y^2 + \sum_{i=1}^m 3\beta_{7+n+i}(y - \eta_i)^2 \cdot I(y - \eta_i \geq 0) \right. \\
&+ \beta_{11+2n+m}z + 2\beta_{12+2n+m}(y \times z) \\
&\left. + 3\beta_{13+2n+m}(y^2 \times z) + \sum_{i=1}^m 3\beta_{13+2n+m+i}((y - \eta_i)^2 \times z) \cdot I(y - \eta_i \geq 0) \right).
\end{aligned}$$

Kuna ka siin  $e^{s(x,y,z)} > 0$ , siis hüpoteeside (11) kontrollimiseks on uurijatel oluline teada, kas teine tegur on võrdne nulliga.

Olgu ühe suvaliselt fikseeritud vanuse väärtus  $\tau$ , aasta väärtus  $\omega$  ja olgu tegemist naistega (ehk  $z = 1$ ). Tuletiste põhjal olgu järgmised vektorid:

$$\lambda_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \cdot \tau \\ 3 \cdot \tau^2 \\ \vdots \\ 3 \cdot (\tau - \xi_n)^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 2 \cdot \omega \\ 3 \cdot \omega^2 \\ \vdots \\ 3 \cdot (\omega - \eta_m)^2 \end{pmatrix}$$

Valemis (11) toodud nullhüpotees omandab mudelist (10) tunnuse vanus järgi tuletist võttes seega kuju  $H_0: \hat{\beta}'\lambda_1 = 0$ . Samast mudelist Aasta järgi tuletist võttes omandab aga valemis (11) toodud nullhüpotees kuju  $H_0: \hat{\beta}'\lambda_2 = 0$ . Saadud hüpoteese võib Poissoni regressiooni korral kontrollida Waldi testi abil ja kvaasi-Poissoni regressiooni korral  $t$ -testi abil kasutades rakendustarkvara R funktsiooni `estimable` [27].

Waldi testiga leitakse  $\chi^2$ -statistik, mis on antud hüpoteeside korral leitav valemiga:

$$\chi_{DF}^2 = \frac{(\hat{\beta}'\lambda_i)^2}{D(\hat{\beta}'\lambda_i)} = \frac{(\hat{\beta}'\lambda_i)^2}{D(\lambda_i'\hat{\beta})} = \frac{(\hat{\beta}'\lambda_i)^2}{\lambda_i' \cdot D(\hat{\beta}) \cdot \lambda_i}, \quad i = 1,2,$$

kus  $\hat{\beta}$  on arvatav valemist (4) ja  $DF$  on vabadusastmete arv, mis sõltub murdepunktide arvust ja kas mudelis kasutatakse sootunnust või mitte [27][29].

Kvaasi-Poissoni jaotust otseselt ei eksisteeri, vaid see on modifikatsioon Poissoni jaotusest (vt ptk 2.1.4). Seetõttu eeldatakse, et parameetrite hinnangud on asümptootiliselt normaaljaotusega, kuna  $n \geq 30$  [20]. Seetõttu on statistilise olulisuse arvutamiseks kasutatud  $t$ -testi. Teststatistik  $t$  on arvatav valemiga:

$$t_{DF} = \frac{\hat{\beta}'\lambda_i}{\sqrt{D(\hat{\beta}'\lambda_i)}} = \frac{\hat{\beta}'\lambda_i}{\sqrt{D(\lambda_i'\hat{\beta})}} = \frac{\hat{\beta}'\lambda_i}{\sqrt{\lambda_i' \cdot D(\hat{\beta}) \cdot \lambda_i}}, \quad i = 1,2,$$

kus  $\hat{\beta}$  on samuti arvatav valemist (4) ja  $DF$  on vabadusastmete arv, mis sõltub murdepunktide arvust ja kas mudelis kasutatakse sootunnust või mitte [20][27].

Kasutades  $\chi^2$ -jaotuse või  $t$ -jaotuse kvantiile on välja arvatud ka olulisuse tõenäosus ehk  $p$ -väärtus. Käänupunktide ning kasvamis- ja kahanemiskohtade tuvastamiseks on kasutatud järgnevat reeglit ( $i = 1,2$ ):

$p$ -väärtus $> \alpha$	$\Rightarrow$ Tegemist on käänukohaga
$p$ -väärtus $< \alpha$	$\nearrow \hat{\beta}'\lambda_i > 0 \quad \Rightarrow$ Tegemist on kasvamiskohaga
	$\searrow \hat{\beta}'\lambda_i < 0 \quad \Rightarrow$ Tegemist on kahanemiskohaga

Kuna mõne vähipaikme korral hakkavad vähijuhud esinema ainult alates teatud vanusest (vt jooniseid 2, 8 ja 11), siis esimeste vanuste korral on antud piirkonda vaadeldud kui käänukohta, kus ei esine ei kasvamis- ega kahanemist riskis.

Käesolevas magistritöös on mudeli käänukohtade ning kasvamis- ja kahanemiskohtade tuvastamiseks kasutatud rakendustarkvara R funktsiooni:

`estimable(mudel, lambda)` (vt lisa 5).

## 2.5 Binoomtest

Käesolevas magistritöös soovitakse hinnata tõenäosust ehk riski haigestuda olemasolevate andmete põhjal. Lisaks sellele leitakse vaadeldavale suurusele ka usaldusintervall. [12]

Olgu  $X_1, X_2, \dots, X_n$  sama jaotusega sõltumatud juhuslikud suurused, mis on genereeritud Bernoulli jaotusest  $B(1, p)$  ja mille tõenäosusfunktsioon on:

$$P(X_i = k) = \begin{cases} 1 - p, & \text{kui } k = 0 \\ p, & \text{kui } k = 1 \end{cases}$$

Sellisel juhul on suurus  $Y = \sum_{i=1}^n X_i$  binoomjaotusega  $B(n, p)$  ja selle tõenäosusfunktsioon on:

$$P(Y = k) = C_n^k p^k (1 - p)^{n-k}, \quad \text{kus } k = 0, 1, \dots, n. \quad [12][20]$$

Parameetri  $p$  hinnang on avaldatav kujul:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{Y}{n}. \quad [12]$$

Antud magistritöös on leitud Clopper–Pearson'i täpsed usaldusintervallid. Need on esitatavad kujul:

$$\left\{ \theta \mid P[B(n, \theta) \leq y] \geq \frac{\alpha}{2} \right\} \cap \left\{ \theta \mid P[B(n, \theta) \geq y] \geq \frac{\alpha}{2} \right\},$$

kus  $y$  on tulemuste arv valimis ja  $B(n, \theta)$  on binoomjaotusega juhuslik suurus  $n$  katsega ja tulemuse saamise tõenäosusega  $\theta$ . [33]

Käesolevas magistritöös on testiks kasutatud rakendustarkvara R funktsiooni:

```
test <- binom.test(juhtumitearv, Rahvaarv)
```

```
test$conf.int[1]; test$conf.int[2] (vt lisa 5).
```

## 2.6 Haigestumuskordaja standardiseerimine vanuse ja soo järgi

Rahvastiku vanuseline jaotus on ajas muutunud, kuna rahvastik vananeb ja inimeste eluiga pikeneb [8]. Analüüsi autor otsustas võtta juhtumite arvu prognoosimisel arvesse ka rahvastiku vanuselist ja soolist jaotust. Soovitud on välja selgitada, missugune oleks haigestunute arv mingil kindlal aastal modelleeritud riski järgi, kui vanuseline struktuur oleks olnud sama, mis 2016. aastal.

Antud meetodika selgitamiseks kasutatakse ühte lihtsat näidet. Olgu rahvastik jaotatud vanuse järgi kolme rühma: noored, keskealised ja vanemaalised. Olgu rahvastiku vanuseline ja sooline jaotus ning mudeli põhjal riskid vähki haigestuda 1968. aastal ja 2016. aastal:

**Tabel 2.** Rahvastiku soolise ja vanuselise osakaalude näide

Vanus	Osakaal kogu rahvastikust		Sugu	Osakaal kogu rahvastikust ( $p_i$ (aasta))		Haigestunuid mudeli põhjal 1968. aastal 10 000 inimese kohta ( $\pi_i$ )	1968. aastal haigestunute hinnangu täpsus ehk dispersioon ( $s_i^2$ )
	1968	2016		1968	2016		
Noored	50%	30%	Mees	25%	14%	10	2.1
			Naine	25%	16%	8	1.7
Keskealised	30%	40%	Mees	13%	18%	43	5.6
			Naine	17%	22%	27	3.8
Vanemaealised	20%	30%	Mees	8%	10%	88	12.5
			Naine	12%	20%	65	9.9

Seega keskmine haigestunute arv  $\gamma_{1968}$  1968. aastal on arvutatav valemiga:

$$\gamma_{1968} = \sum_{i=1}^6 \pi_i \cdot p_i(1968).$$

Antud näite korral on selle suuruse väärtuseks:

$$\gamma_{1968} = 10 \cdot 0.25 + 8 \cdot 0.25 + 43 \cdot 0.13 + 27 \cdot 0.17 + 88 \cdot 0.08 + 65 \cdot 0.12 = 29.52.$$

Kui aga 1968. aastal rahvastiku sooline ja vanuseline jaotus oleks sama olnud, mis 2016. aastal, siis keskmine haigestunute arv  $\gamma'_{1968}$  on arvutatav valemiga:

$$\gamma'_{1968} = \sum_{i=1}^6 \pi_i \cdot p_i(2016).$$

Antud näite korral on selle suuruse väärtuseks:

$$\gamma'_{1968} = 10 \cdot 0.14 + 8 \cdot 0.16 + 43 \cdot 0.18 + 27 \cdot 0.22 + 88 \cdot 0.10 + 65 \cdot 0.20 = 38.16.$$

Dispersioon vaadeldavale suurusele on hinnatav valemiga:

$$\begin{aligned} s_{\gamma'_{1968}}^2 &= D(\gamma'_{1968}) = D\left(\sum_{i=1}^6 \pi_i \cdot p_i(2016)\right) = \sum_{i=1}^6 D(\pi_i \cdot p_i(2016)) \\ &= \sum_{i=1}^6 D(\pi_i) \cdot (p_i(2016))^2 = \sum_{i=1}^6 s_i^2 \cdot (p_i(2016))^2 \end{aligned} \quad (12)$$

Võib eeldada, et valem (12) kehtib, sest tegemist on reaaleluliste andmetega, kus ühe inimgrupi haigestumiste arv ei tohiks mõjutada teise inimgrupi haigestumiste arvu.

Antud näite korral on vaadeldava hinnangu väärtuseks:

$$\begin{aligned} s_{\gamma'_{1968}}^2 &= 2.1 \cdot 0.14^2 + 1.7 \cdot 0.16^2 + 5.6 \cdot 0.18^2 + 3.8 \cdot 0.22^2 + 12.5 \cdot 0.10^2 + 9.9 \cdot 0.20^2 \\ &\approx 0.97. \end{aligned}$$

Usaldusintervall vaadeldavale suurusele on arvutatav valemiga:

$$I_{95\%} = \left( \gamma'_{1968} - z_{\alpha} \cdot \sqrt{s_{\gamma'_{1968}}^2}; \gamma'_{1968} + z_{\alpha} \cdot \sqrt{s_{\gamma'_{1968}}^2} \right) [20].$$

Antud näite korral on vaadeldava suuruse väärtuseks:

$$I_{95\%} = (38.16 - 1.96 \cdot 0.97; 38.16 + 1.96 \cdot 0.97) \approx (36.26; 40.06). \quad (13)$$



Käesolevas magistritöös on rahvastiku jaotust vaadatud soo järgi ja eraldi kõiki vanuseid „0“, „1“, ..., „84“ ja „85+“. Kirjeldatud näitega samal põhimõttel on standardiseeritud vähki haigestumiste arvu 1968.–2015. aastal arvestades 2016. aasta rahvastiku vanuselist ja soolist jaotust.

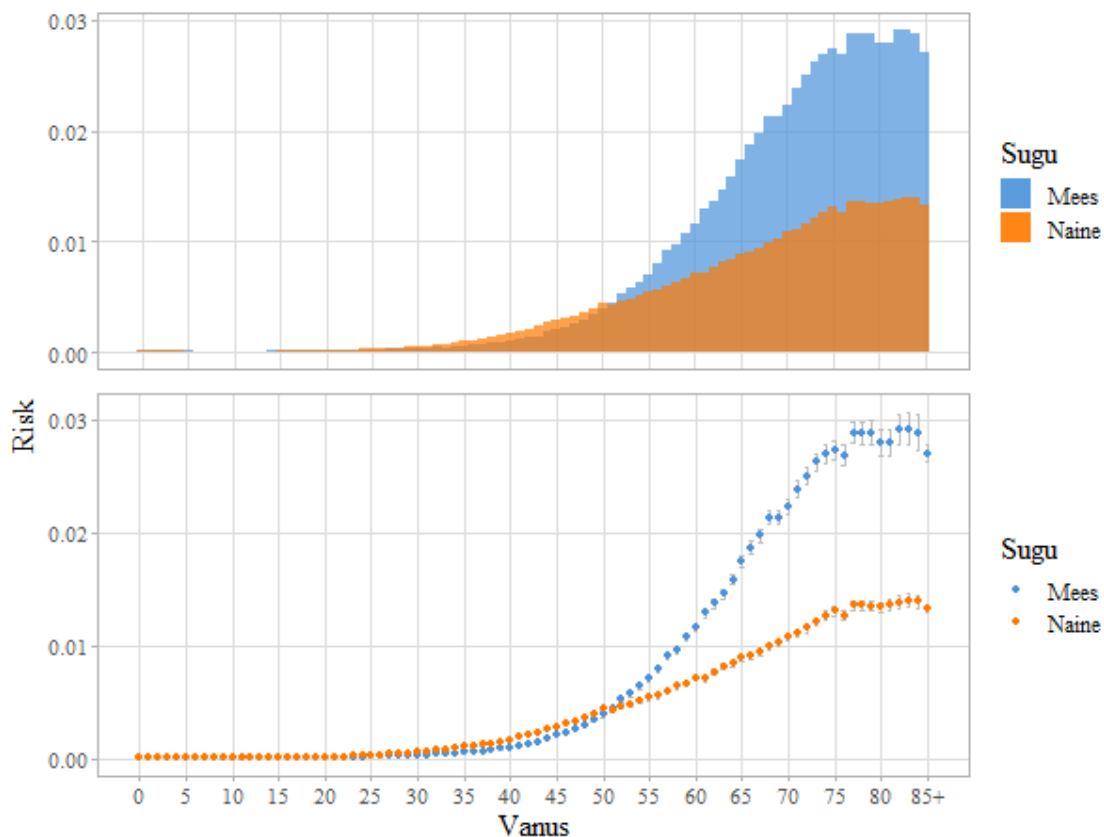
Vähijuhtumite arvu standardiseerimise R kood rahvastiku soo ja vanusjaotuse järgi on välja toodud lisa 6.

## 3 Kirjeldav analüüs

Kirjeldava analüüsi osas uuritakse riski haigestuda summaarselt vanuse ja diagnoosimise aasta järgi ning kõikide paikmete korral ka soo järgi. Lisaks sellele on vaadatud vähijuhtude arvu vanuse järgi aastal 2000 ja diagnoosimise aasta järgi 60aastaste patsientide puhul ning võimalusel eraldi soo järgi.

### 3.1 Kõik paikmed

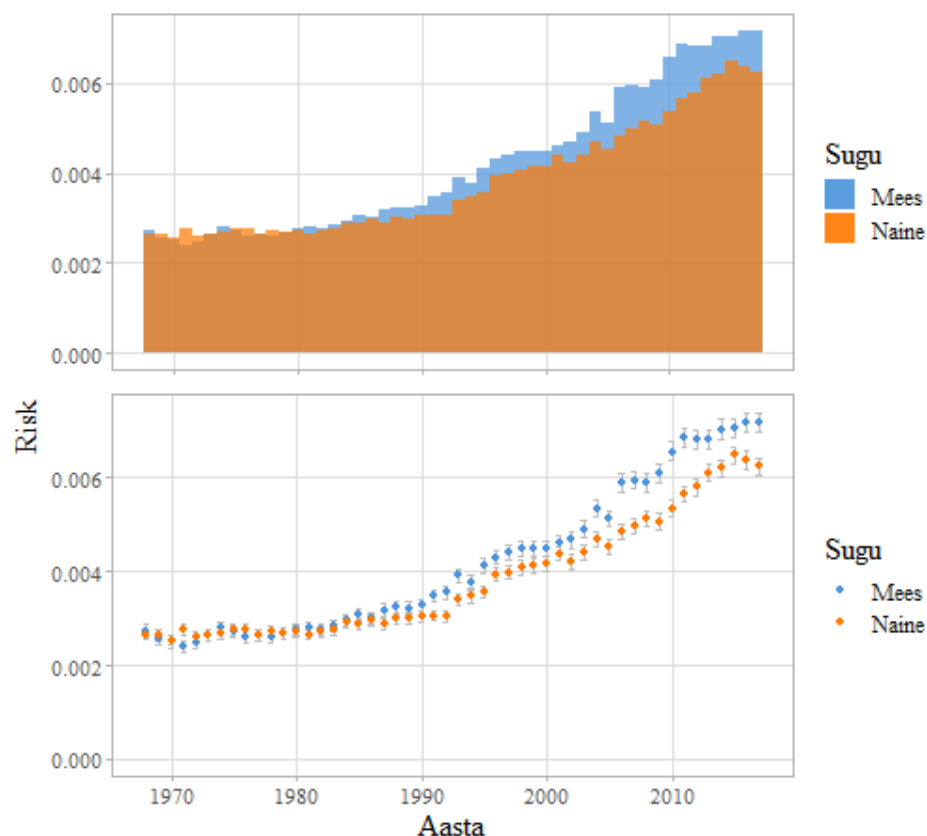
Joonisel 3 on välja toodud üleüldine vähki haigestumise risk vanuse ja soo järgi. Samuti on välja toodud usaldusintervallid, mis on leitud kasutades R-i *binom.test* funktsiooni.



**Joonis 3.** Vähki haigestumise risk vanuse ja soo järgi kõikide paikmete korral

Jooniselt 3 on näha, et risk haigestuda vähki varases eas on väike ja hakkab keskeas kasvama. Vanusevahemikus 30–50 on naistel suurem risk haigestuda vähki ja umbes 50-st eluaastast alates on see näitaja meestel suurem. Mõlema soorühma korral on risk vähki haigestuda kõige suurem alates vanusest 75. Meestel on vanemas eas risk vähki haigestuda peaaegu kaks korda suurem kui naistel. 95%-usaldusintervallide põhjal on näha, et tegemist on ka statistiliselt olulise erinevusega. Võib järeldada, et vanematel inimestel on suurem risk vähki haigestuda.

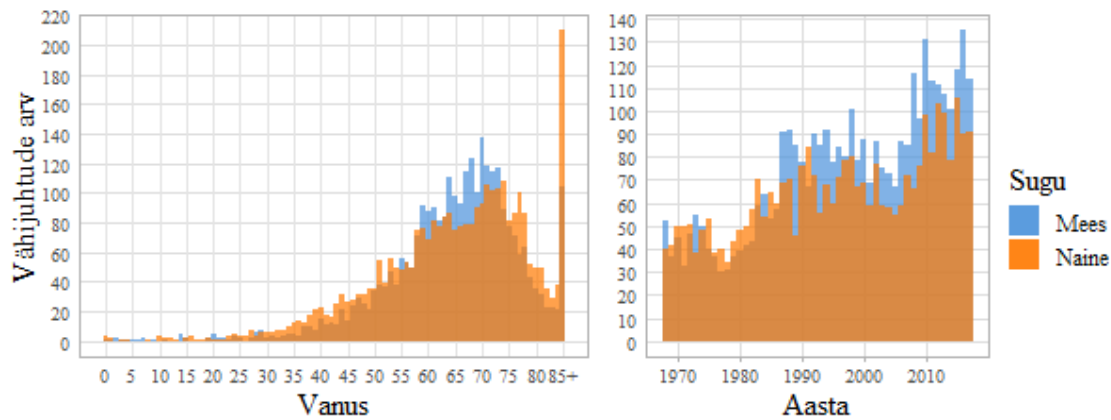
Joonisel 4 on välja toodud üleüldine vähki haigestumise risk diagnoosimise aasta ja soo järgi. Samuti on välja toodud usaldusintervallid, mis on leitud kasutades R-i *binom.test* funktsiooni.



**Joonis 4.** Vähki haigestumise risk vanuse ja soo järgi kõikide paikmete korral, 1968–2017

Jooniselt 4 näeme, et aastate jooksul on mõlema soorühma korral vähki haigestumise risk kasvanud. Naiste puhul on see risk kasvanud veidi vähem kui meeste puhul – erinevus on hakanud tekkima 80ndate aastate keskpaigas. 95%-usaldusintervallide põhjal võib öelda, et tegemist on statistiliselt olulise erinevusega. Uurija arvates võib seda põhjustada peamiselt naiste ja meeste erinev elustiil – näiteks erinev toitumine [16]. Jooniselt on veel näha, et risk vähki haigestuda on 50 aasta vältel suurenenud peaaegu kaks korda. Seda võib põhjustada rahvastiku vananemine, industriaalühiskonna areng ja inimeste tervisekäitumise muutumine aastate jooksul, näiteks ebatervislik toitumine, ülekaal jms [7][8][16].

Joonisel 5 on välja toodud 2000. aastal meeste ja naiste vähki haigestumiste arv vanuse järgi ning 60aastaste vähki haigestumiste arv diagnoosimise aasta järgi kõikide vähipaikmete korral.

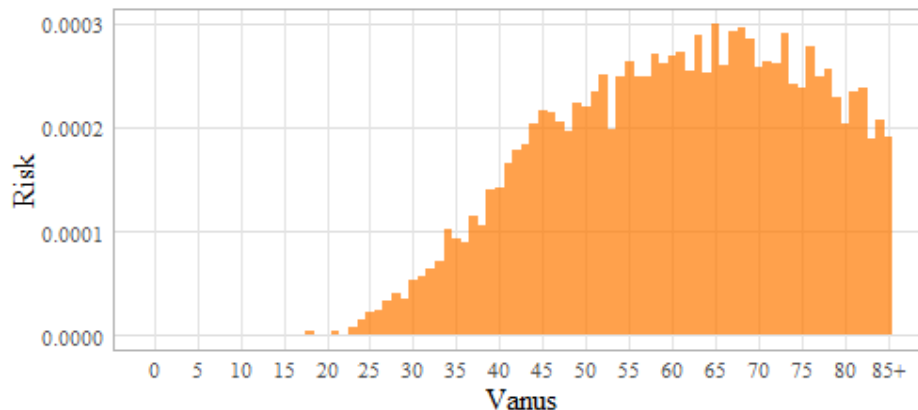


**Joonis 5.** Meeste ja naiste vähki haigestumiste arv vanuse järgi 2000. aastal ning 60aastaste vähki haigestumiste arv diagnoosimise aasta järgi kõikide vähipaikmete korral

Jooniselt 5 näeme, et 2000. aastal oli kõikide paikmete korral haigestumiste arv vanuse järgi kui ka vähijuhtude arv 60aastaste hulgas diagnoosimise aasta järgi väärtuselt suured. See on oluline, sest üksikjuhtumid ei tohiks loodavat mudelit tugevalt mõjutada. Uuriija eeldab, et vaadeldavatele andmetele on võimalik mudelit luua nii, et see oleks ka praktikas rakendatav.

### 3.2 Emakakaelavähk

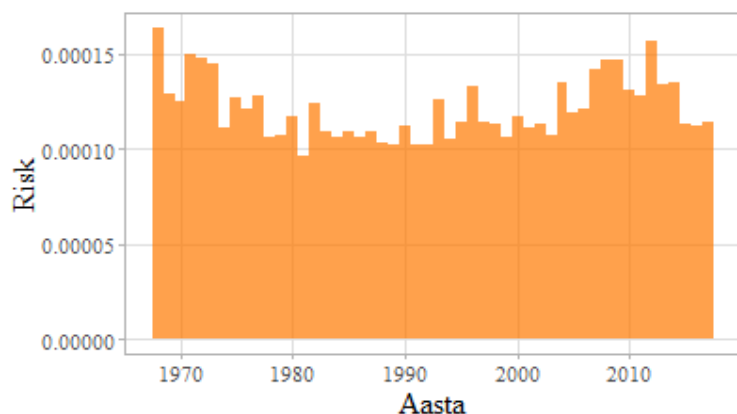
Joonisel 6 on välja toodud emakakaelavähki haigestumise risk vanuse järgi. Antud vähk esineb ainult naistel (vt ptk 1.1.2.2).



**Joonis 6.** Emakakaelavähki haigestumise risk naise vanuse järgi

Jooniselt 6 näeme, et risk haigestuda emakakaelavähki hakkab kasvama juba 20ndates eluaastates ja kasvab eriti kiiresti vanusevahemikus 25–45 eluaastat. Haigestumise risk on eriti kõrge vanusevahemikus 44–84 eluaastat, kus see jääb valdavalt vahemikku 0.0002-0.0003. Risk on langevas trendis eakamate hulgas. Uuriija arvates võib eelkirjeldatud suundumuse põhjuseks olla inimeste aktiivsem seksuaalelu ja sellega alustamine 20ndates eluaastates, mis võib olla ka emakakaelavähi üheks riskifaktoriks (vt ptk 1.1.2.2).

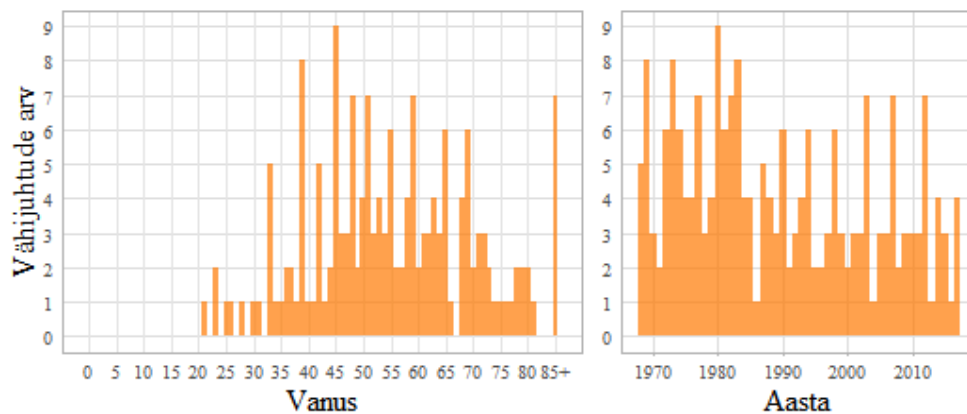
Joonisel 7 on välja toodud naiste emakaelavähki haigestumise risk diagnoosimise aasta järgi.



**Joonis 7.** Emakakaelavähki haigestumise risk diagnoosimise aasta järgi, 1968–2017

Jooniselt 7 on näha, et emakakaelavähki haigestumise risk on olnud aastate jooksul pidevalt muutumises ja mingeid üldiseid kasvamise või kahanemise trende ei saa välja tuua.

Joonisel 8 on välja toodud 2000. aastal emakakaelavähki haigestumiste arv naise vanuse järgi ning 60aastaste naiste emakakaelavähki haigestumiste arv diagnoosimise aasta järgi.

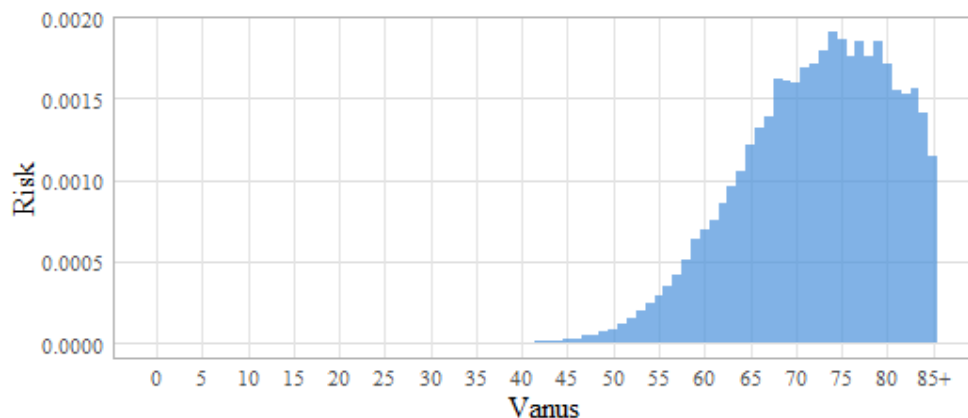


**Joonis 8.** Emakakaelavähki haigestumiste arv naise vanuse järgi 2000. aastal ja 60aastaste emakakaelavähki haigestumiste arv diagnoosimise aasta järgi

Jooniselt 8 on näha, et 2000. aastal oli emakakaelavähki haigestumiste arv naise vanuse järgi kui ka samal aastal 60aastaste hulgas diagnoosimise aasta järgi väga väike. Antud joonise põhjal tuleb selgelt esile, et üksikjuhtumid hakkavad loodavat mudelit väga tugevalt mõjutama. Uuriija eksperthinnangu põhjal ei pruugi olla võimalik luua vaadeldavate andmetele mudelit nii, et antud detailsusastmega prognoosid oleksid praktikas ka rakendatavad.

### 3.3 Eesnäärmevähk

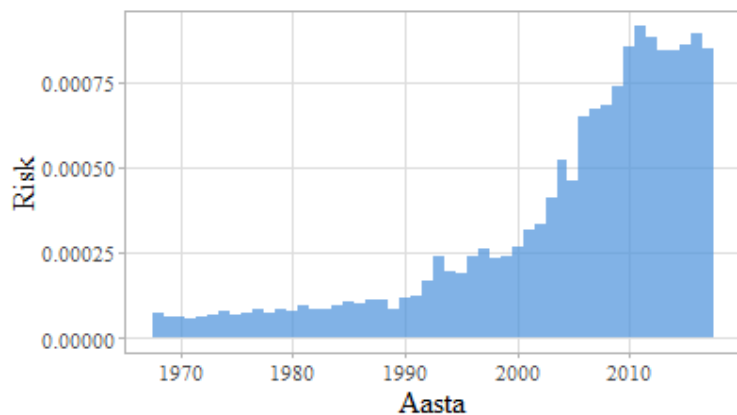
Joonisel 9 on välja toodud eesnäärmevähki haigestumise risk vanuse järgi. Antud vähk esineb ainult meestel (vt ptk 1.1.2.3).



**Joonis 9.** Eesnäärmevähki haigestumise risk mehe vanuse järgi

Jooniselt 9 näeme, et risk haigestuda eesnäärmevähki muutub märkimisväärseks meestel alates vanusest 40. Nooremas eas meestel on haigestumise risk antud vähki peaaegu olematu. Kõige kõrgem eesnäärmevähki haigestumise risk on vanusevahemikus 68–83 eluaastat, kus see ulatub vahemikku 0.0015-0.0020, mis on uurija arvates ka loogiline, sest kõrge vanus on ka üks eesnäärmevähi riskifaktoritest (vt ptk 1.1.2.3).

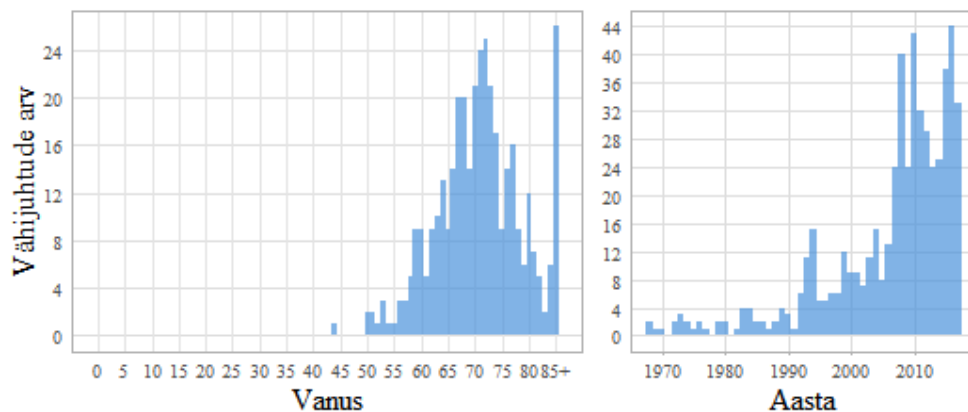
Joonisel 10 on välja toodud meeste eesnäärmevähki haigestumise risk diagnoosimise aasta järgi.



**Joonis 10.** Eesnäärmevähki haigestumise risk diagnoosimise aasta järgi, 1968–2017

Jooniselt 10 näeme, et risk haigestuda eesnäärmevähki on viimase 50 aasta jooksul mitmekordselt kasvanud. Kõige drastilisem haigestumise riski tõus oli perioodil 2000–2010. Uurija arvates võib põhjuseks olla see, et rahvastik vananeb, mistõttu on eesnäärmevähki diagnoositud rohkem, kuna üheks riskiteguriks on mehe kõrge vanus [8]. Analüüsi läbiviija arvates võib riski mõjutada ka meeste ebatervislikud toitumisharjumused ja ülekaalulisus [16][17].

Joonisel 11 on välja toodud 2000. aastal eesnäärmevähki haigestumiste arv mehe vanuse järgi ja eesnäärmevähki haigestumiste arv 60aastaste meeste hulgas diagnoosimise aasta järgi.



**Joonis 11.** Eesnäärmevähki haigestumiste arv 2000. aastal mehe vanuse järgi ja eesnäärmevähki haigestumiste arv 60aastaste meeste hulgas diagnoosimise aasta järgi

Jooniselt 11 näeme, et 2000. aastal oli eesnäärmevähki haigestumiste arv mehe vanuse järgi kui ka 60aastaste meeste haigestumiste arv diagnoosimise aasta järgi üsna väike. Üksikjuhtumid võivad uurija hinnangul loodavat mudelit mõjutada, kuid mitte nii tugevalt kui näiteks emakaelavähi puhul.

## 4 Mudelite kirjeldus ja prognoosivõime

Vähki haigestumise riski vaadeldakse kõikide vähipaikmete korral eraldi soo järgi. Emakakaelavähki haigestumise riski on võimalik vaadelda ainult naistel ja eesnäärmevähi korral ainult meestel. Selleks, et saada ülevaade aasta mõju kohta vähki haigestumise riskile, on uuritud riski muutumist aastate vältel 60aastaste inimeste hulgas. Vanuse mõju uurimiseks on vaadeldud riski muutumist vanuse järgi aastal 2000. Mudelid on hinnatud 1968.–2016. aasta andmete põhjal ja on kasutatud kõigis vanustes inimeste andmeid.

Vanuse või diagnoosimise aasta mõju riskile vaatlemiseks graafikul tuuakse välja esmalt genereeritud mudeli punase joonega ning mustade täppidega tähistatakse tegeliku riski väärtused vastavalt 2000. aastal või 60aastaste hulgas. Mudeli valiku ja genereerimise meetodika on selgitatud peatükkides 2–2.3. Järgmisel joonisel on lisaks eelnevalt nimetatutele välja toodud ka riskide 95%-usaldusintervallid – nende leidmise meetodika on selgitatud peatükis 2.5. Kolmandal joonisel on välja toodud mudeli riski kasvamisepunktid punase täpiga ja kahanemispunktid roheline täpiga – kui tegemist on seost kirjeldava funktsiooni käänukohaga, siis vaadeldava väärtuse korral on see tähistatud musta täpiga. Punktide klassifitseerimise meetodika on selgitatud peatükis 2.4.

Lisaks on uuritud iga paikme mudeli prognoose 2017. aasta vähijuhtumite arvu osas ja võrreldud seda tegelike 2017. aasta vähijuhtumite arvuga (vt R koodi lisas 7). Joonisel on iga riski väärtuse korral leitud 95%-usaldusintervallid – nende leidmise meetodika on selgitatud peatükis 2.5.

### 4.1 Kõik paikmed

Splaini koostades on tunnuse Vanus korral kasutatud  $n = 10$  murdepunkti ja tunnuse Aasta korral kasutatud  $m = 9$  murdepunkti.

Mudeli kuju kõikide paikmete korral on (tähistused on selgitatud osas (8)):

$$\frac{E(y_i)}{n_i} = e^{s(x,y,z)},$$

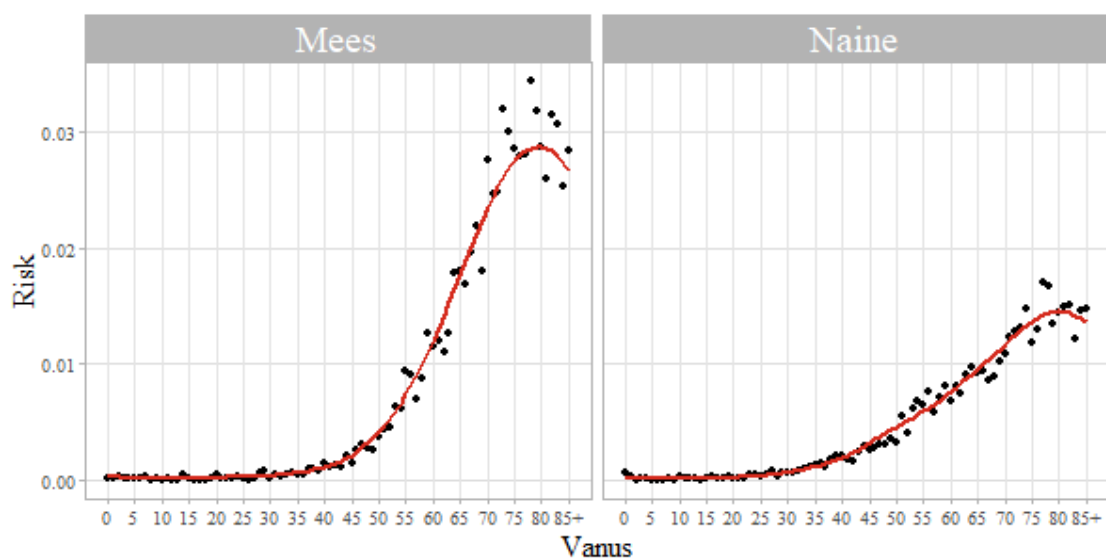
kus  $s(x, y, z)$  kuju on näha lisas 8.

Antud mudeli puhul oli suhte (7) väärtus  $1.473 < 2$ , mistõttu ülehajuvus pole märkimisväärne ja mudelis on kasutatud tavalist Poissoni regressiooni.



### 4.1.1 Vanus

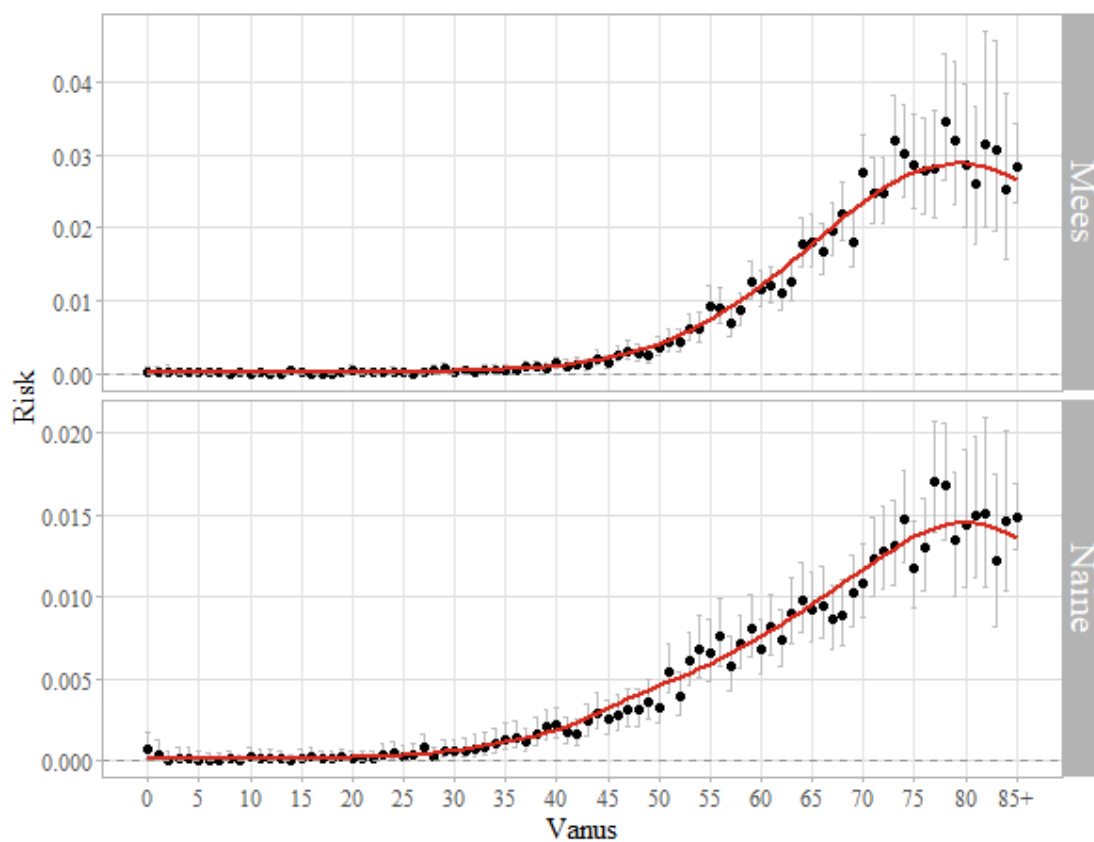
Joonisel 12 on välja toodud mudeli kuju (punane joon) ja tegelikud riski väärtused (mustad täpid) soo ning vanuse järgi kõikide paikmete korral 2000. aastal.



**Joonis 12.** Soo ja vanuse mõju haigestumise riskile kõikide vähipaikmete korral, 2000

Jooniselt 12 on näha, et vähki haigestumise risk on eakatel meestel ligi kaks korda suurem kui eakatel naistel. Üldjoontes on juba antud graafiku põhjal näha, et loodud mudel prognoosib kõigi paikmete korral vanuse järgi riski üsna hästi. Mõlema soo puhul on võimalik välja tuua, et noores eas on vähki haigestumise risk nullilähedane ja keskeas hakkab risk kasvama.

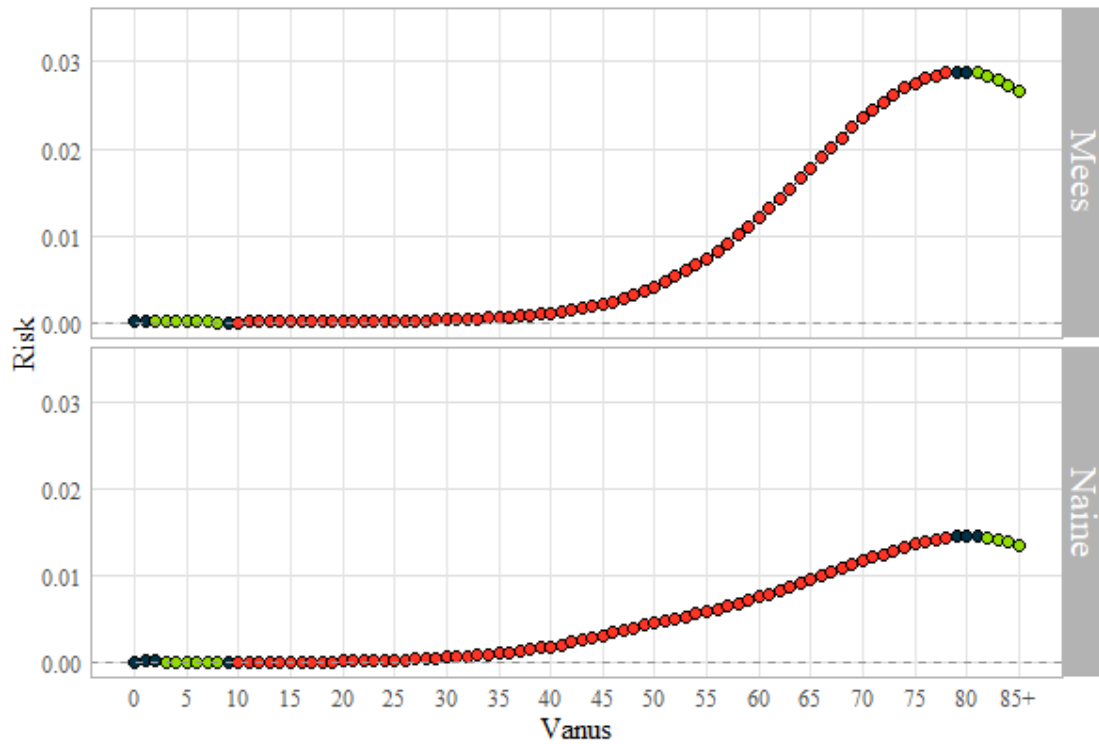
Joonisel 13 on lisaks välja toodud kõikide vanuste jaoks hinnatud vähki haigestumuse riskid koos 95%-usaldusintervalliga.



**Joonis 13.** Soo ja vanuse mõju haigestumise riskile kõikide vähipaikmete korral koos 95%-usaldusintervallidega, 2000

Jooniselt 13 on näha, et nii meeste kui ka naiste vähki haigestumise riski prognoosides käitub mudel ootuspäraselt – see püsib kõigi vanuste korral 95%-usaldusintervallide piires. Antud joonise puhul võib veel välja tuua, et mida kõrgem on vanus, seda laiemad on ka 95%-usaldusintervallid. See aga näitab, et kõrgemas eas on raskem adekvaatselt hinnata vähki haigestumise riski.

Joonisel 14 on välja toodud seost kirjeldava funktsiooni käänu-, kasvamis- ja kahanemiskohad kõikide paikmete korral 2000. aastal.

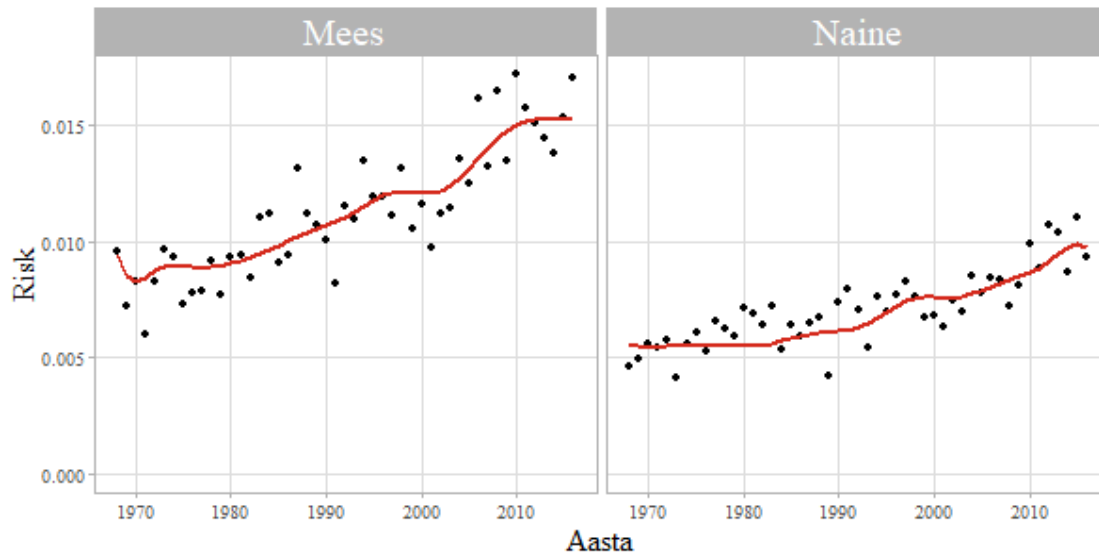


**Joonis 14.** Mudeli kasvamis-, kahanemis- ja käänukohad soo ja vanuse järgi kõikide paikmete korral, 2000

Jooniselt 14 on näha, et kõikide vähipaikmete korral toimub mudelis näitajate kasv nii meeste kui naiste puhul vanusevahemikus 10–78 eluaastat. Noores eas esineb mõlema soo korral seost kirjeldava funktsiooni kahanemine – naistel vanusevahemikus 3–8 ja meestel 2–8 eluaastat, kuid absoluutarvudes ei ole tegemist siiski märkimisväärse muutusega. Kõrges eas esineb taas näitajate kahanemist 81aastaste ja vanemate meeste ning 82aastaste ja vanemate naiste puhul.

#### 4.1.2 Diagnoosimise aasta

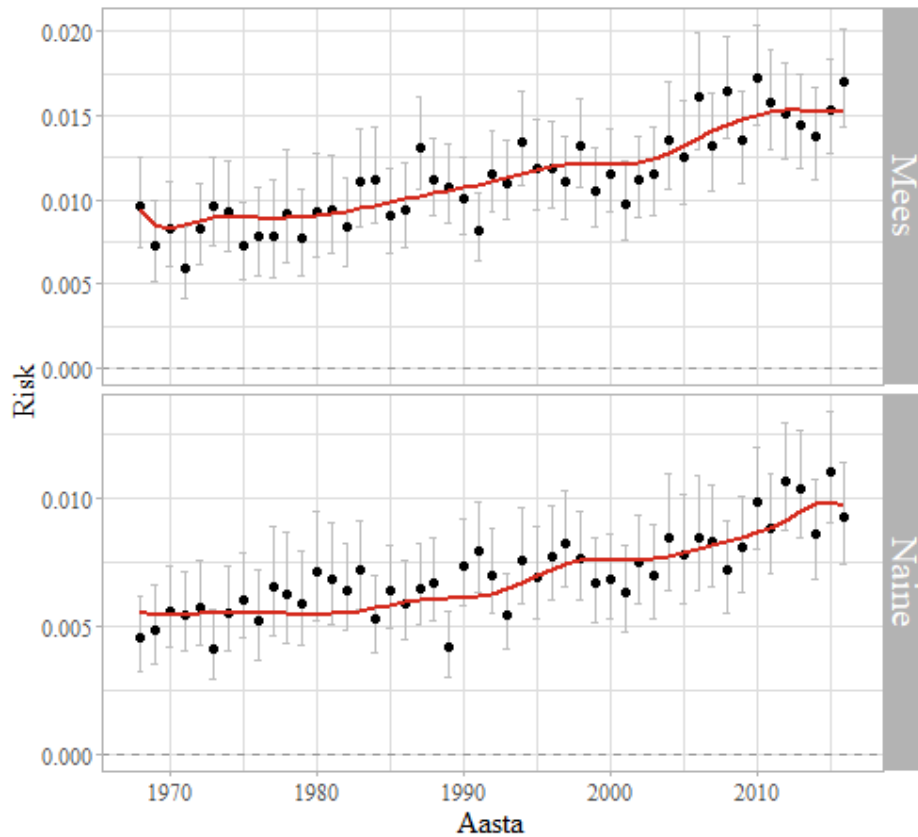
Joonisel 15 on välja toodud mudeli kuju (punane joon) ja tegelikud riski väärtused (mustad täpid) soo ning diagnoosimise aasta järgi kõikide paikmete korral 60aastaste hulgas.



**Joonis 15.** Soo ja diagnoosimise aasta mõju haigestumise riskile kõikide vähipaikmete korral 60aastaste hulgas, 1968–2016

Jooniselt 15 on näha, et mudeli järgi on sama vanadel meestel risk haigestuda vähki olnud läbi aastate kõrgem kui naistel. Nii meestel kui naistel on antud näitaja tõusnud vaadeldavas ajavahemikus, kuid meestel on see olnud palju järsem kui naistel. Samuti võib juba antud jooniselt näha, et risk haigestuda vähki on üsna muutuv sõltuvalt aastast.

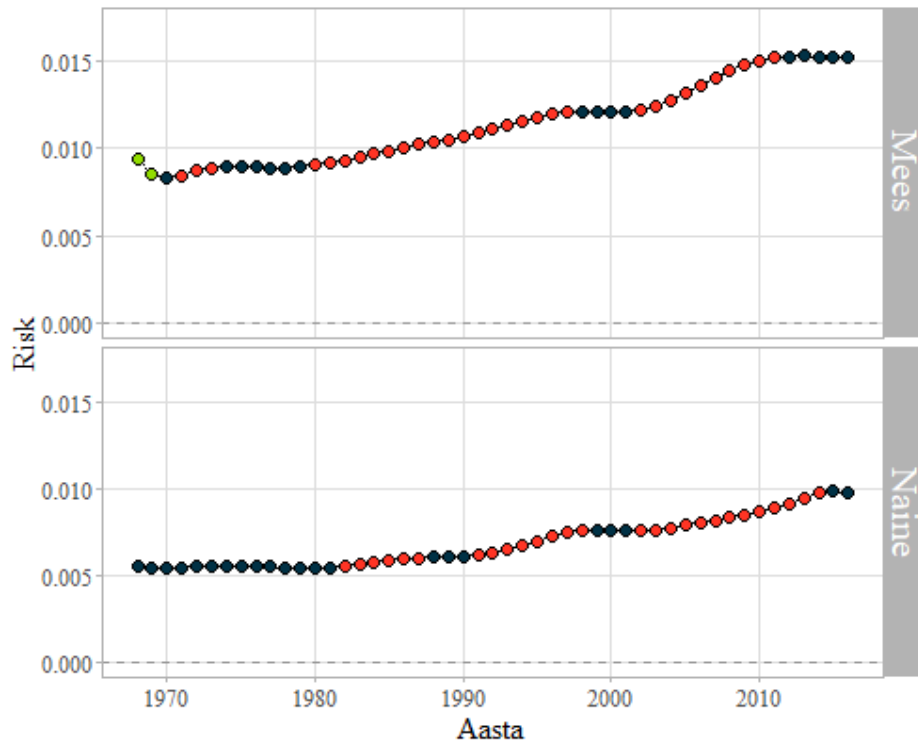
Joonisel 16 on lisaks välja toodud kõikide aastate jaoks hinnatud vähki haigestumuse riskid koos 95%-usaldusintervalliga.



**Joonis 16.** Soo ja diagnoosimise aasta mõju haigestumise riskile kõikide vähipaikmete korral koos 95%-usaldusintervallidega, 1968–2016

Jooniselt 16 on näha, et meeste puhul jääb mudel usaldusintervallidest välja 1991. aastal ja naiste puhul 1989. aastal. Ülejäänud aastatel jäävad mõlema soo näitajad mudelis 95%-usaldusintervallidesse. See on ootuspärane mudeli käitumine, sest tegemist on 95%-usaldusintervallidega ja ootuspäraselt 5% 95%-usaldusintervallidest ei peagi tegelikku haigestumise riski katma. Veel on võimalik jooniselt näha, et aja jooksul on risk haigestuda statistiliselt oluliselt tõusnud, sest varasemate aastate riskide 95%-usaldusintervallid ei löiku hilisemate aastate 95%-usaldusintervallidega.

Joonisel 17 on välja toodud seost kirjeldava funktsiooni käänu-, kasvamis- ja kahanemiskohad kõikide paikmete korral 60aastaste seas.

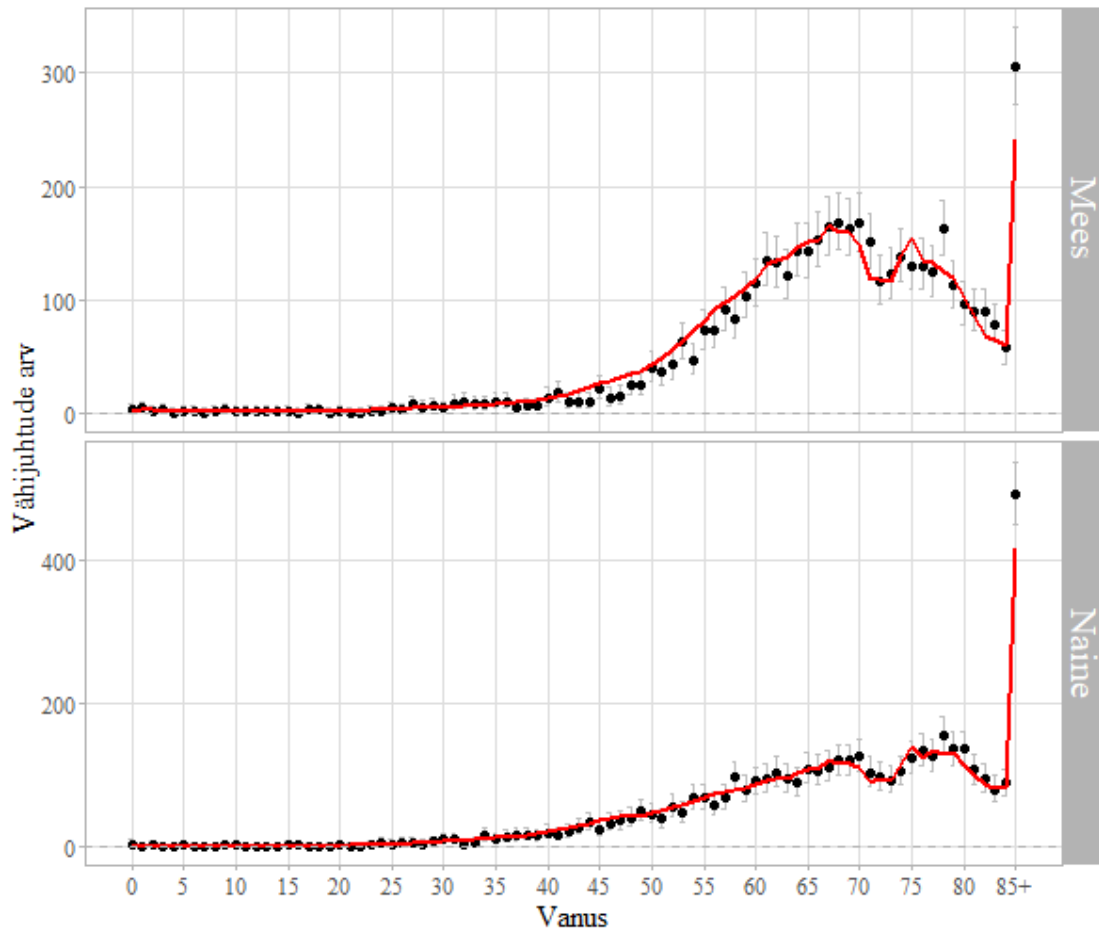


**Joonis 17.** Mudeli kasvamis-, kahanemis- ja käänukohad soo ja diagnoosimise aasta järgi kõikide paikmete korral 60aastaste hulgas, 1968–2016

Jooniselt 17 on näha, et 60aastaste naiste puhul on võimalik märgata mudelis ainult vähki haigestumise riski tõuse, mis esinesid ajavahemikel 1982–1987, 1991–1998 ja 2002–2014. Sama vanade meeste puhul on märgata seost kirjeldaval funktsioonil langusi aastatel 1968–1969 ja tõuse ajavahemikel 1971–1973, 1980–1997 ja 2002–2011.

### 4.1.3 Mudeli prognoosivõime

Joonisel 18 on välja toodud mudeli põhjal leitud vähki haigestumise arvu prognoos 2017. aastaks (punane joon) ja 2017. aasta tegelikud väärtused (mustad täpid) koos usaldusintervallidega (hallid jooned). Vaadeldavaid andmeid on uuritud soo ja vanuse järgi.



**Joonis 18.** Mudeli prognoosivõime 2017. aastaks koos tegelike 2017. aasta andmetega kõikide vähipaikmete korral

Jooniselt 18 on näha, et meeste puhul jääb mudel 95%-usaldusintervalli piiridest välja 78aastaste korral ja mõlema soo puhul 85aastaste ja vanemate korral. Ülejäänud vanuste puhul nii meestel kui naistel satub mudel enamvähem 95%-usaldusintervallidesse. See on ootuspärane mudeli käitumine, sest tegemist on 95%-usaldusintervallidega ja 5% 95%-usaldusintervallidest ei pea tegelikku haigestumisriski katma. Uurijale oli ka ootuspärane, et 85aastaste ja vanemate inimeste haigestumiste arvu on raske korrektselt modelleerida – positiivsena võib välja tuua selle, et mudel siiski liigub tegeliku haigestunute arvu väärtuse suunas ja püüab tegelikku väärtust saavutada.

## 4.2 Emakakaelavähk

Emakakaelavähi haigestumise riskimudeli splaini koostades on tunnuse vanus korral kasutatud  $n = 2$  murdepunkti ja tunnuse Aasta korral kasutatud  $m = 3$  murdepunkti.

Mudeli kuju kõikide paikmete korral on (tähistused on selgitatud osas (8)):

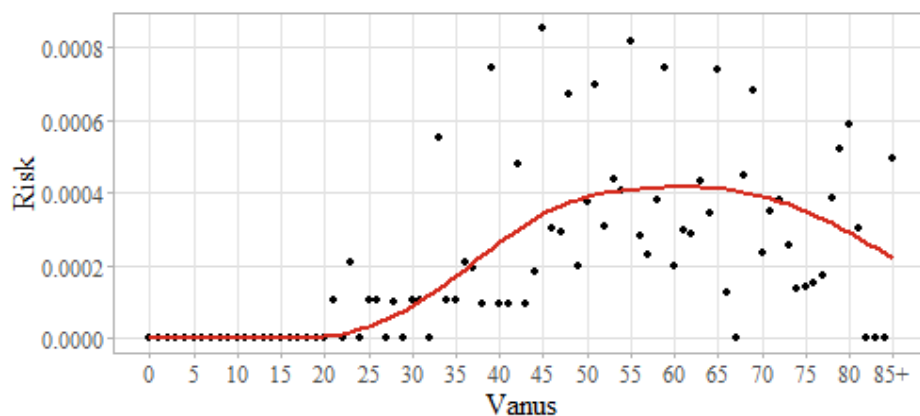
$$\frac{E(y_i)}{n_i} = e^{s(x,y,z)},$$

kus  $s(x, y, z)$  kuju on näha lisas 9.

Antud mudeli puhul oli suhte (7) väärtus  $1.041 < 2$ , mistõttu ülehajuvust ei esine ja mudelis on kasutatud tavalist Poissoni regressiooni.

## 4.2.1 Vanus

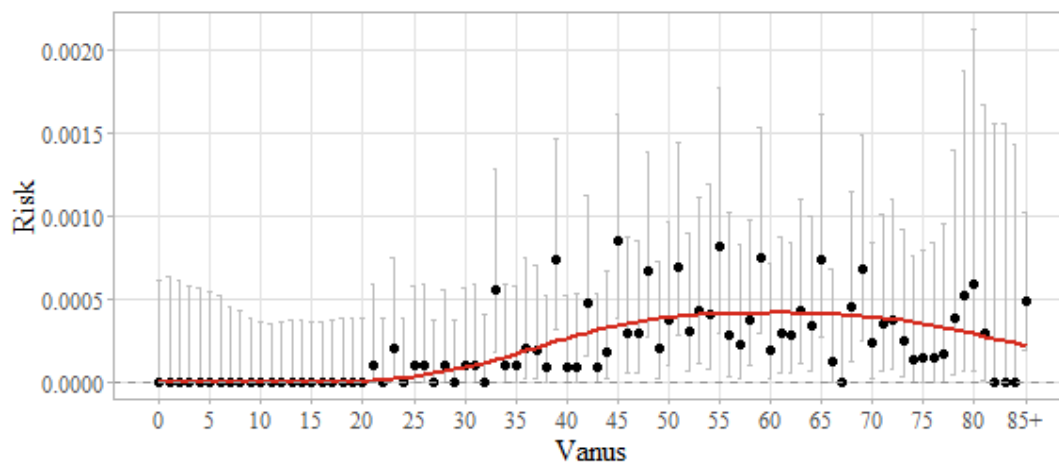
Joonisel 19 on välja toodud emakakaelavähi riski mudeli kuju (punane joon) ja tegelikud riski väärtused (mustad täpid) vanuse järgi 2000. aastal.



**Joonis 19.** Vanuse mõju haigestumise riskile emakakaelavähi korral, 2000

Jooniselt 19 on näha, et haigestumise risk on suuresti muutuv sõltuvalt naise vanusest. Sealhulgas tuleb ka välja tuua, et emakakaelavähi korral on haigestumise riski väärtus üsna väike, mistõttu võib eeldada, et üksikjuhtumid võivad tugevalt mõjutada loodavat mudelit (vt joonist 8).

Joonisel 20 on lisaks välja toodud kõikide vanuste jaoks hinnatud vähki haigestumuse riskid koos 95%-usaldusintervalliga.

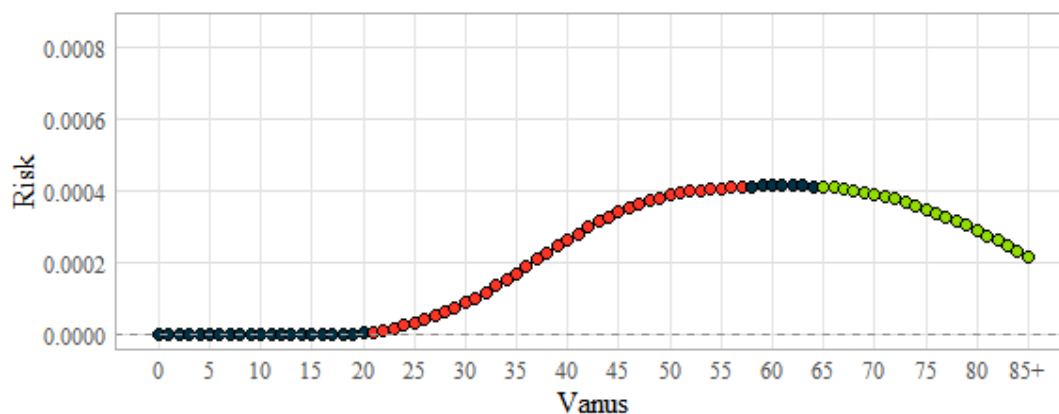


**Joonis 20.** Vanuse mõju haigestumise riskile emakakaelavähi korral koos 95%-usaldusintervallidega, 2000



Jooniselt 20 on näha, emakakaelavähki haigestumise riski prognoosides käitub mudel ootuspäraselt – see jääb peaaegu kõigi vanuste korral 95%-usaldusintervalli piiridesse, välja arvatud 33aastaste ja 45aastaste puhul. Kuni 5%-l juhtudel esinevad vead on ootuspärased 95%-usaldusintervallide puhul. Antud joonise puhul võib veel välja tuua, et kõikide vanuste korral on riskil väga laiad 95%-usaldusintervallid. Põhjus võib olla selles, et emakakaelavähi juhtumeid esinebki vähe (vt joonist 8). Lisaks on väga raske leida statistiliselt olulisi erinevusi riskis erinevate vanuste vahel, kuna nende usaldusintervallid enamjaolt lõikuvad omavahel. See aga näitab, et emakakaelavähki haigestumise riski jaoks on väga raske leida ühte konkreetset mudelit – väga palju mudeleid sobiks kirjeldama emakakaelavähki haigestumise riski.

Joonisel 21 on välja toodud seost kirjeldava funktsiooni käänu-, kasvamis- ja kahanemiskohad emakakaelavähi korral 2000. aastal.

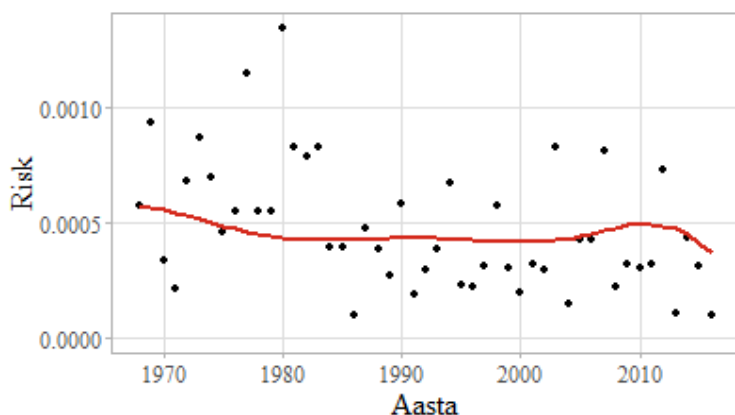


**Joonis 21.** Mudeli kasvamis-, kahanemis- ja käänukohad naise vanuse järgi emakakaelavähi korral, 2000

Jooniselt 21 on näha, et emakakaelavähi jaoks loodud mudeli järgi leiab riskinäitajate kasv aset vanusevahemikus 21–57 eluaastat. Seost kirjeldavas funktsioonis esineb kahanemist kõrgemas eas ehk 65aastaste ja vanemate naiste puhul.

#### 4.2.2 Diagnoosimise aasta

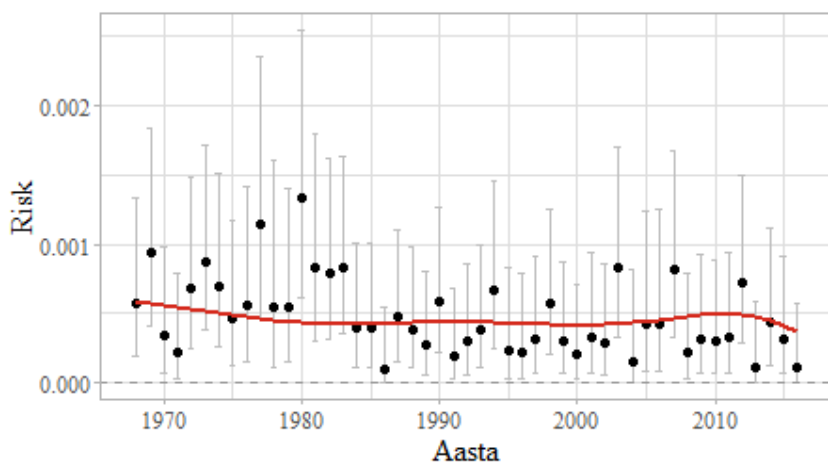
Joonisel 22 on välja toodud mudeli kuju (punane joon) ja tegelikud riski väärtused (mustad täpid) diagnoosimise aasta järgi emakakaelavähi korral 60aastaste naiste hulgas.



**Joonis 22.** Diagnoosimise aasta mõju haigestumise riskile emakakaelavähi korral 60aastaste naiste hulgas, 1968–2016

Jooniselt 22 on näha, et mudeli järgi on risk haigestuda emakakaelavähki 60aastaste naiste puhul püsinud läbi aastate umbes samal tasemel. Samuti võib juba siit jooniselt näha, et risk haigestuda emakakaelavähki on vägagi muutuv sõltuvalt aastast – põhjus võib olla taaskord see, et emakakaelavähi juhtumite arv on väike ja mudel on tugevalt mõjutatud üksikjuhtumitest (vt joonist 8).

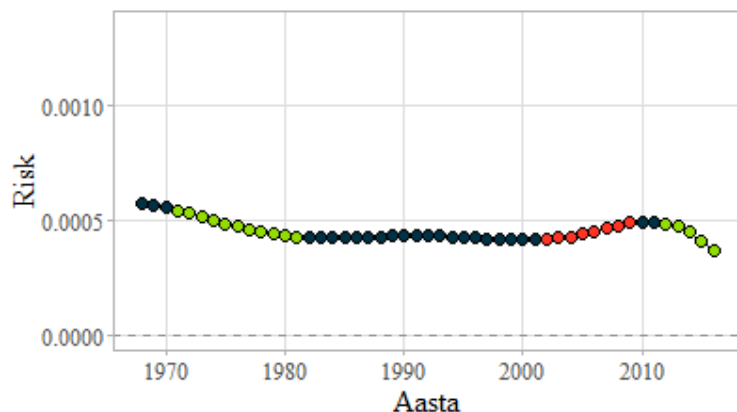
Joonisel 23 on lisaks välja toodud kõikide aastate jaoks hinnatud vähki haigestumuse riskid koos 95%-usaldusintervalliga.



**Joonis 23.** Diagnoosimise aasta mõju haigestumise riskile emakakaelavähi korral koos 95%-usaldusintervallidega, 1968–2016

Jooniselt 23 on näha, et peaaegu kõikide diagnoosimise aastate puhul satub mudel 95%-usaldusintervalli piiridesse, v.a aastal 1980. See on taaskord ootuspärane mudeli käitumine, sest umbes 5%-l juhtudest võib esineda vigu 95%-usaldusintervallide puhul. Veel on võimalik jooniselt näha, et 95%-usaldusintervallid on väga laiad ning väga palju mudeleid sobiks kirjeldama emakakaelavähi haigestumise riski. See aga näitab taas, et emakakaelavähi haigestumise riski jaoks on väga raske leida ühte konkreetset mudelit. Samuti on ka emakakaelavähi korral väga raske leida statistiliselt olulisi erinevusi aastate vahel, kuna nende usaldusintervallid suures osas lõikuvad.

Joonisel 24 on välja toodud seost kirjeldava funktsiooni käänu-, kasvamis- ja kahanemiskohad emakakaelavähi korral 60aastaste naiste seas.

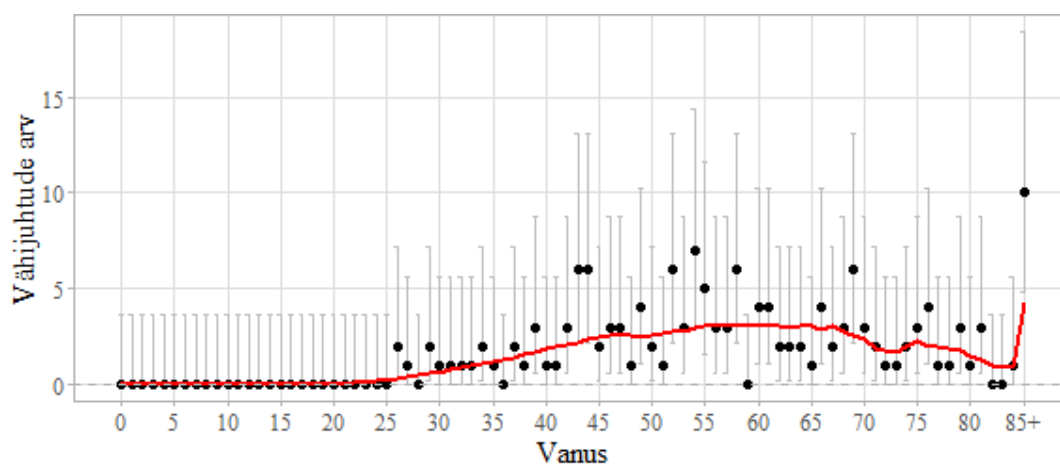


**Joonis 24.** Mudeli kasvamis-, kahanemis- ja käänukohad diagnoosimise aasta järgi emakakaelavähi korral 60aastaste naiste hulgas, 1968–2016

Jooniselt 24 on näha, et 60aastaste naiste puhul on võimalik märgata seost kirjeldavas funktsioonis emakakaelavähi haigestumise riski kasvamist ajavahemikul 2002–2009 ja kahanemist ajavahemikel 1971–1981 ja 2012–2016. Ülejäänud aastatel on tegemist käänukohtadega.

### 4.2.3 Mudeli prognoosivõime

Joonisel 25 on välja toodud mudeli põhjal leitud emakakaelavähi haigestunumise arvu prognoos 2017. aastaks (punane joon) ja võrreldud seda 2017. aasta tegelike väärtustega (mustad täpid) koos usaldusintervallidega (hallid jooned). Vaadeldavaid andmeid on uuritud naiste vanuse järgi.



**Joonis 25.** Mudeli prognoosivõime 2017. aastaks koos tegelike 2017. aasta andmetega emakakaelavähi korral

Jooniselt 25 on näha, et mudel jääb 95%-usaldusintervallidest välja 85aastaste ja vanemate naiste puhul. See on uurija arvates ka loogiline, sest prognoosides kõrges vanuses inimeste haigestunute arvu on väga raske saada adekvaatseid tulemusi. Tuleb ka tähele panna, et 2017. aastal on

emakakaelavähi korral riskidel väga laiad 95%-usaldusintervallid, mistõttu sobib antud suurust kirjeldama väga palju erinevaid mudeleid ja mudelid on tugevalt mõjutatud üksikjuhtumitest.

Emakakaelavähki haigestumise riski analüüsi põhjal saab uurija teha järelduse, et väheste juhtumite arvuga vähipaikmete korral on mudelid tugevalt mõjutatud üksikjuhtumitest. Seetõttu antud detailsusastmega prognooside tegemine ei ole uurija arvates mõistlik.

## 4.3 Eesnäärmevähk

Eesnäärmevähki haigestumise riskimudeli splaini koostades on tunnuse Vanus korral kasutatud  $n = 1$  murdepunkt ja tunnuse Aasta korral kasutatud  $m = 2$  murdepunkti.

Mudeli kuju kõikide paikmete korral on (tähistused on selgitatud osas (8)):

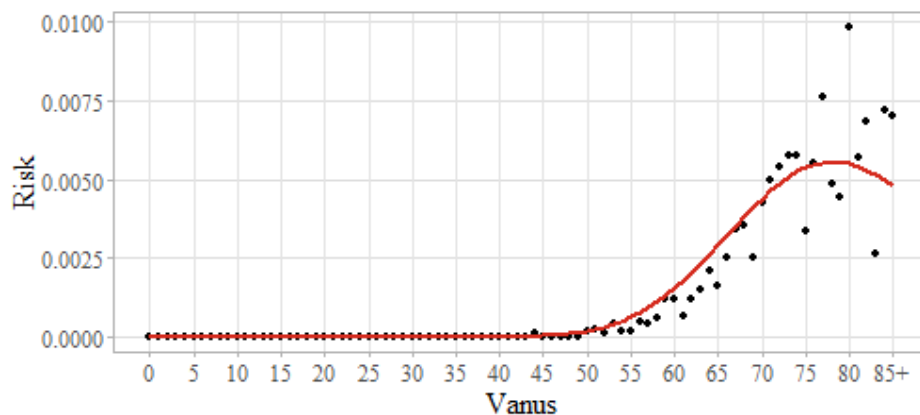
$$\frac{E(y_i)}{n_i} = e^{s(x,y,z)},$$

kus  $s(x, y, z)$  kuju on näha lisas 10.

Antud mudeli puhul oli suhte (7) väärtus  $2.158 > 2$ , mistõttu esineb ülehajuvus ja mudelis on kasutatud kvaasi-Poissoni regressiooni.

### 4.3.1 Vanus

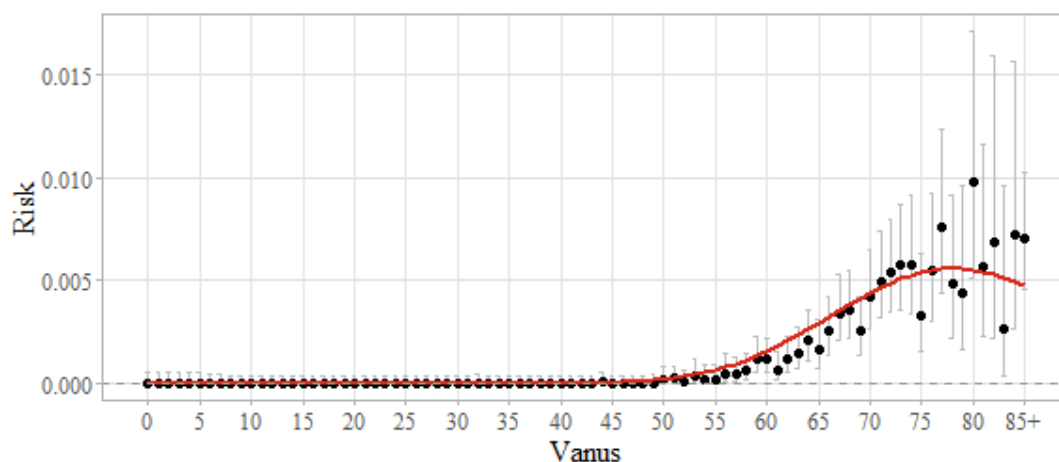
Joonisel 26 on välja toodud mudeli kuju (punane joon) ja tegelikud riski väärtused (mustad täpid) vanuse järgi eesnäärmevähi korral 2000. aastal.



**Joonis 26.** Vanuse mõju haigestumise riskile eesnäärmevähi korral, 2000

Jooniselt 26 on näha, et risk haigestuda eesnäärmevähi on nooremas eas meestel nullilähedane ja alates umbes 50ndast eluaastast hakkab risk drastiliselt kasvama. Kõrgemas eas on riski väärtus tugevas muutumises sõltuvalt vanusest. Üldjoontes on juba antud graafiku põhjal näha, et loodud mudel prognoosib vanuse järgi üsna hästi riski haigestuda eesnäärmevähi.

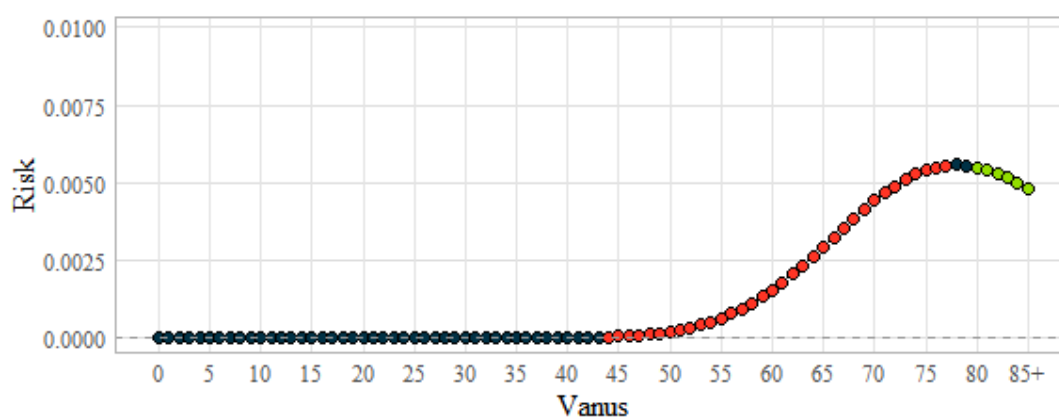
Joonisel 27 on lisaks välja toodud kõikide vanuste jaoks hinnatud vähki haigestumuse riskid koos 95%-usaldusintervalliga.



**Joonis 27.** Vanuse mõju haigestumise riskile eesnäärmevähi korral koos 95%-usaldusintervallidega, 2000

Jooniselt 27 on näha, et meeste eesnäärmevähi haigestumise riski prognoosides käitub mudel ootuspäraselt – see on kõigi vanuste korral 95%-usaldusintervalli piirides. Mida kõrgem on vanus, seda laiemad on ka 95%-usaldusintervallid. See aga näitab, et kõrgemas eas ongi raskem adekvaatselt hinnata vähki haigestumise riski. Joonise põhjal on siiski võimalik täheldada teatud statistiliselt olulisi erinevusi haigestumise riskis nooremate ja vanemate meeste vahel – näiteks ei löiku usaldusintervallid 62aastaste ja 72aastaste meeste puhul ehk 72aastastel meestel oli statistiliselt oluliselt suurem risk haigestuda eesnäärmevähi, kui 62aastastel meestel 2000. aastal. Selliseid statistiliselt olulisi erinevusi on võimalik täheldada veel väga paljude vanusepaaride korral.

Joonisel 28 on välja toodud seost kirjeldava funktsiooni käänu-, kasvamis- ja kahanemiskohad eesnäärmevähi korral 2000. aastal.



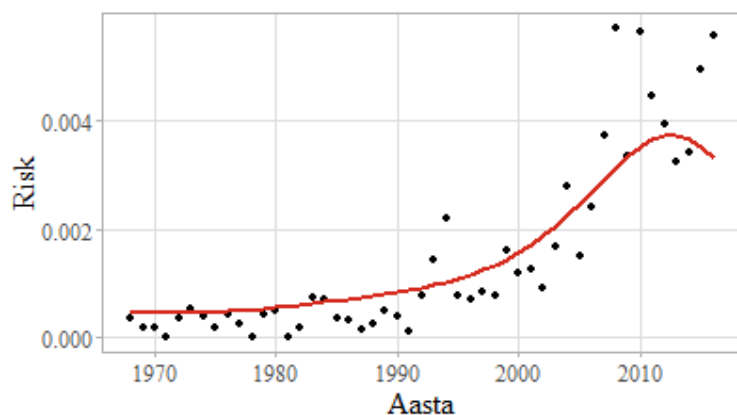
**Joonis 28.** Mudeli kasvamis-, kahanemis- ja käänukohad mehe vanuse järgi eesnäärmevähi korral, 2000

Jooniselt 28 on näha, et eesnäärmevähi korral on seost kirjeldav funktsioon kasvavas trendis meestel vanusevahemikus 18–77 eluaastat. Riski väärtust vaadeldes on näha, et selle

märkimisväärne kasvamine esineb alates vanusest 44. Mudeli põhjal esineb haigestumisriski kahanemist taas kõrges eas ehk 80aastaste ja vanemate meeste puhul.

### 4.3.2 Diagnoosimise aasta

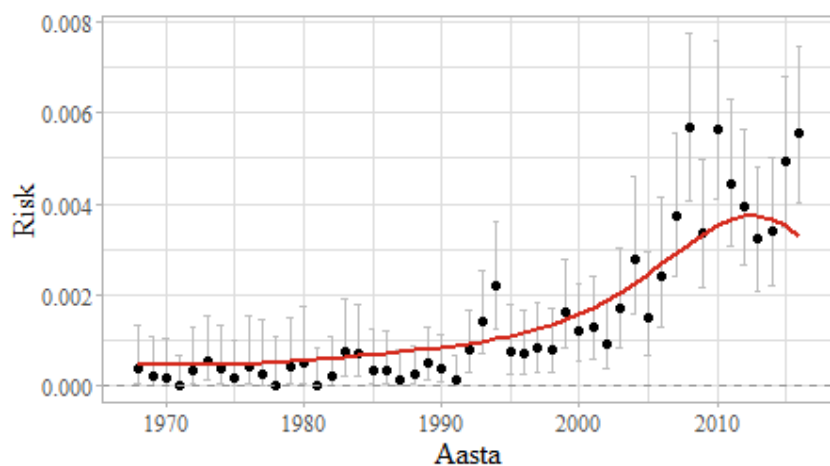
Joonisel 29 on välja toodud mudeli kuju (punane joon) ja tegelikud riski väärtused (mustad täpid) diagnoosimise aasta järgi eesnäärmevähi korral 60aastaste meeste hulgas.



**Joonis 29.** Diagnoosimise aasta mõju haigestumise riskile eesnäärmevähi korral 60aastaste meeste hulgas, 1968–2016

Jooniselt 29 on näha, et mudeli järgi on 60aastaste meeste puhul risk haigestuda eesnäärmevähki läbi aastate järsult kasvanud. Jooniselt võib juba näha, et risk haigestuda eesnäärmevähki on üsna muutuv sõltuvalt diagnoosimise aastast – eriti ajavahemikus 2005–2016. Samuti võib tähele panna huvitavat erandit aastal 1994.

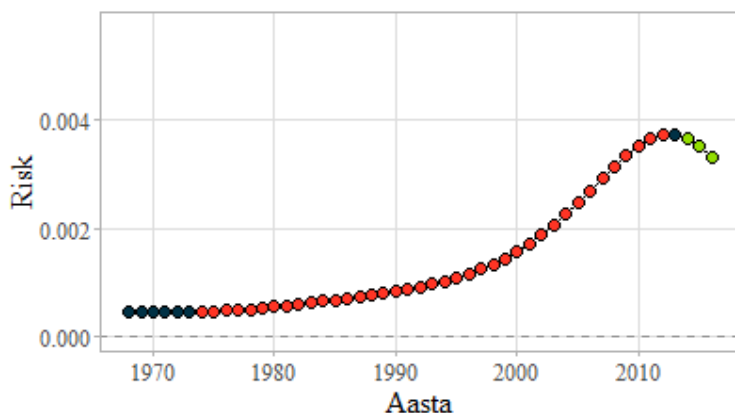
Joonisel 30 on lisaks välja toodud kõikide aastate jaoks hinnatud vähki haigestumuse riskid koos 95%-usaldusintervalliga.



**Joonis 30.** Diagnoosimise aasta mõju haigestumise riskile eesnäärmevähi korral koos 95%-usaldusintervallidega, 1968–2016

Joonisel 30 on näha, et 1994., 2008., 2010. ja 2017. aastal diagnoosi saanud meestel jääb mudel usaldusintervallidest välja. Ülejäänud aastate korral jääb mudel 95%-usaldusintervalli piiridesse. See ei ole väga ootuspärane mudeli käitumine, sest võimalikud vead võivad esineda umbes 5%-l juhtudest – antud juhul ei satu mudel usaldusintervallidesse umbes 8%-l aastatest. Uurija saab selle põhjal teha järelduse, et loodud mudel ei suuda hästi kirjeldada eesnäärmevähi haigestumise riski umbes viimase 10 vaadeldava aasta jooksul. Veel on võimalik antud jooniselt näha, et aja jooksul on risk haigestuda eesnäärmevähi statistiliselt oluliselt tõusnud, sest varasemate aastate riskide 95%-usaldusintervallid ei löiku hilisemate aastate 95%-usaldusintervallidega.

Joonisel 31 on välja toodud seost kirjeldava funktsiooni käänu-, kasvamis- ja kahanemiskohad eesnäärmevähi korral 60aastaste meeste seas.

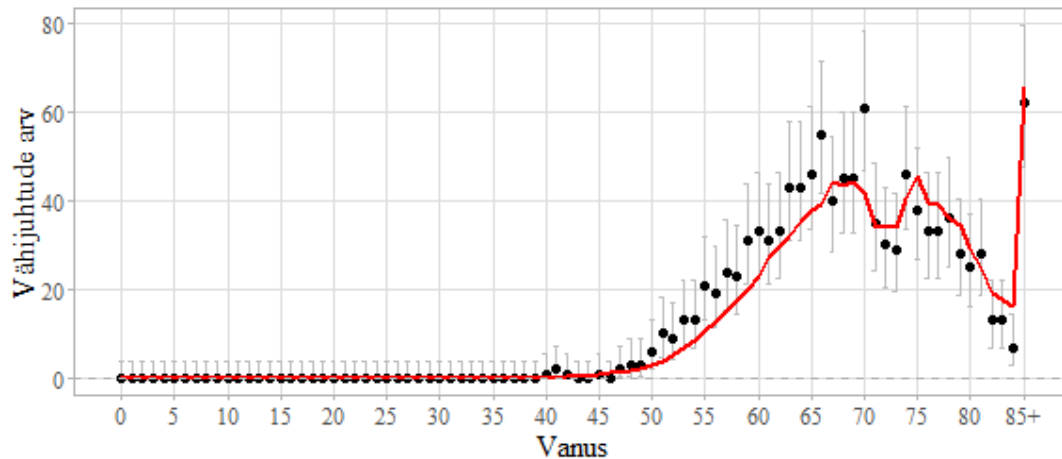


**Joonis 31.** Mudeli kasvamis-, kahanemis- ja käänukohad diagnoosimise aasta järgi eesnäärmevähi korral 60aastaste meeste hulgas, 1968–2016

Joonisel 31 on näha, et 60aastaste meeste puhul on võimalik seost kirjeldavas funktsioonis märgata eesnäärmevähi haigestumise riski kasvu ajavahemikul 1974–2012 ja kahanemist ajavahemikul 2014–2016.

### 4.3.3 Mudeli prognoosivõime

Joonisel 32 on välja toodud mudeli põhjal leitud eesnäärmevähi haigestunumise arvu prognoos 2017. aastaks (punane joon) ja võrreldud seda 2017. aasta tegelike väärtustega (mustad täpid) koos usaldusintervallidega (hallid jooned). Vaadeldavaid andmeid on uuritud meeste vanuse järgi.



**Joonis 32.** Mudeli prognoosivõime 2017. aastaks koos tegelike 2017. aasta andmetega eesnäärmevähi korral

Jooniselt 32 on näha, et mudel alahindab veidi vähki haigestunute arvu vanusevahemikus 50–66, kuid mudel jääb enam-vähem 95%-usaldusintervalli piiridesse või intervallide vahetusse lähedusse. Näiteks 70aastaste ja 84aastaste meeste puhul ei satu küll mudel uuritava tunnuse usaldusintervalli piiridesse, kuid uurija hinnangul ei ole ka tekkinud viga märkimisväärselt suur. Mudel hindab üllatavalt hästi 85aastaste ja vanemate vanusegrupi meeste haigestumise riski 2017. aastal.

Kuna mudel satub valdavalt riskide 95%-usaldusintervallidesse, saab järeldada, et mudeli prognoos on piisavalt täpne ning kogu erinevus hinnangute ja tegelike riski väärtuste vahel võib tekkida vanusepõhise hinnangu ebatäpsusest.

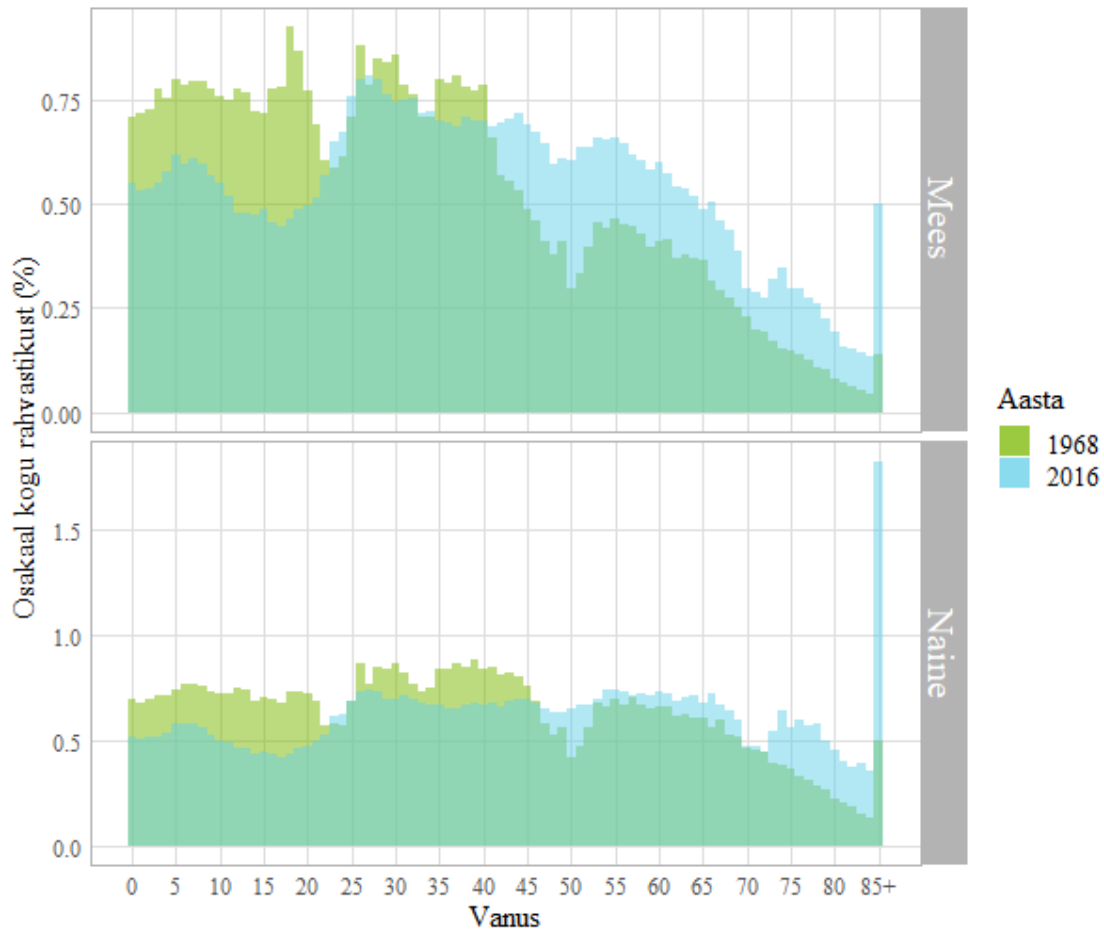
Antud analüüsi põhjal saab uurija teha järelduse, et suurema juhtumite arvuga vähiliigi puhul on võimalik luua selliseid mudeleid, millega saab ka detailsemalt haigestumise riski muutumist prognoosida. Küll aga tuleks uurija arvates olla mõneti ettevaatlik eesnäärmevähi haigestumise riski prognoosimisel eakatele ja viimasel aastakümnel.

## 4.4 Haigestunute arvu standardiseerimine rahvastiku soo ja vanuse järgi

Haigestunute arvu standardiseerimist on tehtud ainult kõikide paikmete korral. Sama meetodikat saab rakendada kõikidele teistele vähipaikmete andmetele. Kõige pealt on standardiseeritud ainult 1968. aasta vähki haigestunute arvu prognoos 2016. aasta rahvastiku soolise ja vanuselise jaotuse järgi.

Joonisel 33 on välja toodud naiste ja meeste osakaal kogu rahvastikust vanuse järgi aastatel 1968 ja 2016.

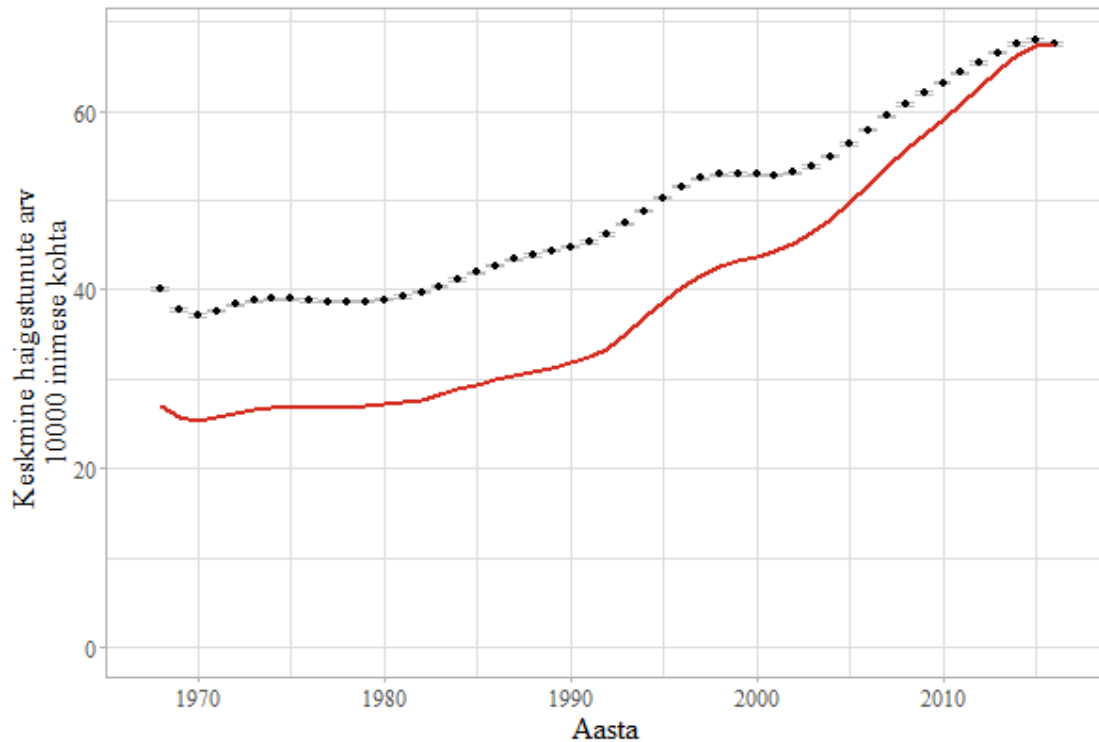




**Joonis 33.** 1968. ja 2016. aasta rahvastiku sooline ja vanuseline jaotus

Jooniselt 33 on näha, et nii vanemate meeste kui ka naiste osakaal rahvastikus on 2016. aastal suurem kui 1968. aastal. Eriti tugevalt on nii meeste kui naiste puhul kasvanud vanuses 85+ inimeste osakaal, mis on uurija arvates ka loogiline, sest rahvastik vananeb ja antud vanusegruppi kuuluvad kõik inimesed, kes on 85aastased ja vanemad, samal ajal, kui muud vanused on joonisel välja toodud ühe kaupa [8].

Joonisel 34 on välja toodud 2016. aasta rahvastiku soolise ja vanuselise jaotuse järgi standardiseeritud keskmised haigestunute arvud (mustad täpid) koos usaldusintervallidega (hallid jooned) ja tegelikud vähki haigestunute arvud (punane joon). Kuna 95%-usaldusintervallid osutusid väga kitsaks, siis neile allolevas analüüsis tähelepanu ei pöörata.



**Joonis 34.** Ajavahemiku 1968–2015 keskmise vähki haigestunute arvu prognoosi standardiseeritud suurused 2016. aasta rahvastiku soolise ja vanuselise jaotuse järgi 10 000 inimese kohta

Jooniselt 34 on näha, et kui rahvastiku soo- ja vanusjaotus aastatel 1968–2011 oleks olnud sama, mis aastal 2016, siis keskmine vähki haigestunute arv oleks olnud statistiliselt oluliselt suurem kui see tegelikult oli. Umbes keskmiselt 10 inimese võrra 10 000 inimese kohta oleks haigestunuid olnud rohkem ajavahemikul 1968–1992, peale mida hakkas vaadeldav vahe kahanema. Uurižale pakkus see fakt huvi, sest see muutus võib olla seotud Eesti taasiseseisvumisega 1991. aastal ja selle järel muutunud eluoluga.

Haigestumuse tõus on osaliselt tingitud elanikkonna vananemisest. Vaatamata standardiseeritud haigestumuskordaja laugemast tõustust, mis on joonisel näha, võib siiski väita, et vähki haigestumine on sagenenud, sest ka standardiseeritud haigestumuskordaja väärtus on aastatega kasvanud.

## Kasutatud kirjandus

- [1] Bogovski, P. (1996). *RHK-10: Rahvusvaheline haiguste ja nendega seotud terviseprobleemide statistiline klassifikatsioon*. Tallinn: Tallinna Raamatutrükikoda.
- [2] Chao S., Chippendale R., jt (2014). *Current Diagnosis & Treatment: Geriatrics, Second Edition: Benign Prostatic Hyperplasia & Prostate Cancer*. McGraw-Hill: New York. Kasutatud 29.03.2020.  
<http://accessmedicine.mhmedical.com.ezproxy.utlib.ut.ee/content.aspx?bookid=953&sectionid=53375664>
- [3] de Jong, P., Heller, G. Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press: United Kingdom.
- [4] Eesti Statistikaamet. *Andmebaasist* (01.10.2010). Kasutatud 05.03.2020.  
<http://pub.stat.ee/px-web.2001/dialog/AAB.htm>
- [5] Eesti vähiliit. *Eesnäärmevähk* (n.d.). Kasutatud 06.03.2020.  
<https://cancer.ee/eesnaarmevahk/>
- [6] Eiche, I.-E.. (2016). *Eesnäärmevähi haigestumus- ja suremustrendid Eestis*. Magistritöö. Tartu Ülikool, peremeditsiini ja rahvatervishoiu instituut.
- [7] Eurostat. *Obesity rate by body mass index (BMI)*. (2020). Kasutatud 25.03.2020.  
[https://ec.europa.eu/eurostat/web/products-datasets/-/sdg\\_02\\_10](https://ec.europa.eu/eurostat/web/products-datasets/-/sdg_02_10)
- [8] Eurostat. *statistics explained: Elanikkonna struktuur ja vananemine*. (2019). Kasutatud 18.03.2020.  
[https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population\\_structure\\_and\\_ageing/et](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing/et)
- [9] Frees, W. E. (2010). *Regression Modeling with Actuarial and Financial Applications*. Cambridge University Press: New York.
- [10] Guichard, D., Farmer, D., Schueller, A. (n.d). *Calculus: early transcendentals: 5. Curve Sketching*. Kasutatud 17.03.2020.  
[https://www.whitman.edu/mathematics/calculus\\_online/chapter05.html](https://www.whitman.edu/mathematics/calculus_online/chapter05.html)
- [11] Hastie, T. J., Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman & Hall: London.
- [12] Heumann, C., Schomaker, M., Shalabh. (2016). *Introduction to Statistics and Data Analysis: With Exercises, Solutions and Applications in R*. Springer.
- [13] Inimene.ee terviseportaal. *Eesnäärmevähk* (n.d.). Kasutatud 06.03.2020.  
<https://www.inimene.ee/haigused-ja-seisundid/list/haigused-ja-seisundid/eesnaarmevahk-327>
- [14] Inimene.ee terviseportaal. *Emakakaelavähk* (n.d.). Kasutatud 06.03.2020.  
<https://www.inimene.ee/haigused-ja-seisundid/list/haigused-ja-seisundid/emakakaela-vahk-333>
- [15] Jacobs C., Robinson L., Webb P. (2014) *Genetics for Health Professionals in Cancer Care: From Principles to Practice*. Oxford University Press.  
DOI: 10.1093/med/9780199672844.001.0001
- [16] Jakobson, M. (2014). *Teavitus kutsub mehi enam puu- ja köögivilju sööma*. Toitumine.ee. Kasutatud 19.03.2020.  
<https://toitumine.ee/artiklid/teavitus-kutsub-mehi-enam-puu-ja-koogivilju-sooma>
- [17] Kensler K. H., Rebbeck T. R. (2020). *Cancer Progress and Priorities: Prostate Cancer*. American Association of Cancer Research. DOI: 10.1158/1055-9965.EPI-19-0412.
- [18] Käärrik, M. *Generalized Linear Models: Lecture 3. Hypothesis testing. Goodness of Fit. Model diagnostics* (2019). Kasutatud 16.03.2020.  
[https://courses.ms.ut.ee/MTMS.01.011/2019\\_spring/uploads/Main/GLM\\_slides\\_3\\_hypothesis\\_gof.pdf](https://courses.ms.ut.ee/MTMS.01.011/2019_spring/uploads/Main/GLM_slides_3_hypothesis_gof.pdf)

- [19] Käärik, M. *Generalized Linear Models: Lecture 7. Count data models I. Poisson model* (2019). Kasutatud (10.03.2020)  
[https://courses.ms.ut.ee/MTMS.01.011/2019\\_spring/uploads/Main/GLM\\_slides\\_7\\_Poisson.pdf](https://courses.ms.ut.ee/MTMS.01.011/2019_spring/uploads/Main/GLM_slides_7_Poisson.pdf)
- [20] Mann, P. S. (2007). *Introductory Statistics: Sixth Edition*. John Wiley & Sons, Inc.
- [21] National Cancer Institute. *Cervical Cancer Prevention (PDQ®)-Patient Version* (2019). Kasutatud 02.04.2020.  
<https://www.cancer.gov/types/cervical/patient/cervical-prevention-pdq>
- [22] National Cancer Institute. *NCI Dictionary of Cancer Terms: carcinoma in situ* (n.d.). Kasutatud 13.04.2020.  
<https://www.cancer.gov/publications/dictionaries/cancer-terms/def/carcinoma-in-situ>
- [23] National Cancer Institute. *NCI Dictionary of Cancer Terms: menorrhagia* (n.d.). Kasutatud 02.04.2020.  
<https://www.cancer.gov/publications/dictionaries/cancer-terms/def/menorrhagia>
- [24] National Cancer Institute. *Prostate Cancer Prevention (PDQ®)-Patient Version* (2019). Kasutatud 02.04.2020.  
<https://www.cancer.gov/types/prostate/patient/prostate-prevention-pdq>
- [25] National Cancer Institute. *What Is Cancer?* (2015). Kasutatud 02.04.2020.  
<https://www.cancer.gov/about-cancer/understanding/what-is-cancer>
- [26] Papadakis M. A., McPhee S. J., Bernstein J. (2020). *Quick Medical Diagnosis & Treatment 2020: Cervical Cancer*. McGraw-Hill: New York. Kasutatud 25.03.2020.  
<http://accessmedicine.mhmedical.com/content.aspx?bookid=2750&sectionid=231365137>
- [27] rdrv.io: *R/estimable.R*. (n.d). Kasutatud 17.03.2020.  
<https://rdrv.io/cran/gmodels/src/R/estimable.R>
- [28] Rencher, A. C., Schaalje G. B. (2008). *Linear Models in Statistics: Second Edition*. John Wiley & Sons: Hoboken, New Jersey.
- [29] Rodríguez, G. *Generalized Linear Models: Tests of Hypotheses*. (2020). Princeton University. Kasutatud 18.03.2020. <https://data.princeton.edu/wws509/notes/a1s2>
- [30] SA Tartu Ülikooli Kliinikumi Hematoloogia-onkoloogia kliinik. *Üldinfo vähktõve kohta* (n.d.). Kasutatud 06.03.2020.  
<https://www.kliinikum.ee/ho/info-haiguste-kohta/83-rinnavaehk>
- [31] Stapleton, H. J. (1995). *Linear Statistical Models*. John Wiley & Sons, Inc.
- [32] Tervise Arengu Instituut. *Vähiregister* (2019). Kasutatud 05.03.2020.  
<https://www.tai.ee/et/tegevused/registrid/vahiregister>
- [33] Tulin, S. (n.d.). *How to Handle Binomial Proportion with Zero Frequency*. Merck & Co., Inc., Whitehouse Station, NJ. Kasutatud 15.04.2020.  
[https://www.lexjansen.com/nesug/nesug13/41\\_Final\\_Paper.pdf](https://www.lexjansen.com/nesug/nesug13/41_Final_Paper.pdf)

## Lisad

### Lisa 1. Pahaloomuliste kasvajate grupid ja esinemiste arv kokku 1968.–2017. aastal

Pahaloomuliste kasvajate grupid RHK-10 järgi	Juhtude arv	Pahaloomuliste kasvajate paikmed detailsemalt	Juhtude arv
Huul, suuõõs, neel (C00–C14)	6 682	Huul (C00)	1 207
		Keel (C01–C02)	1 206
		Ige, suupõhi jm (C03–C06)	1 379
		Suured süljenäärmed (C07–C08)	670
		Mandel, suuneel (C09–C10)	1 114
		Ninaneel (C11)	290
		Pirnsopis, alaneel (C12–C13)	736
		Huul, suuõõs, neel (muu) (C14)	80
Seedeelundid (C15–C26)	72 875	Söögitoru (C15)	2 716
		Magu (C16)	26 131
		Peensool (C17)	581
		Käärsool (C18)	16 452
		Pärasool jm (C19–C21)	11 892
		Maks jm (C22)	3 480
		Sapipõis jm (C23–C24)	2 097
		Kõhunääre (C25)	9 352
		Seedeelundid (muu) (C26)	174
Hingamis-, rindkeresised elundid (C30–C39)	38 555	Ninaõõs, kõrv, ninakõrvalurked (C30–C31)	609
		Kõri (C32)	3 354
		Hingetoru, bronh, kops (C33–C34)	34 180
		Tüümus, süda, keskseinand, pleura (C37–C38)	387
		Hingamiselundid jm (C39)	25
Luu, liigeskõhr (C40–C41)	770		
Naha pahaloomuline melanoom (C43)	5 294		
Naha muud pahaloomulised kasvajad (C44)	31 909		
Mesoteelkude ja pehmed koed (C45–C49)	2 450		
Rind (C50)	24 463		

Eelmise tabeli jätk.

Pahaloomuliste kasvajaite grupid RHK-10 järgi	Juhtude arv	Pahaloomuliste kasvajaite paikmed detailsemalt	Juhtude arv
Naissuguelundid (C51–C58)	26 572	Häbe, tupp (C51–C52)	1 640
		Emakakael (C53)	8 541
		Emakakeha (C54)	8 518
		Emakas (osa täpsustamata) (C55)	69
		Munasari (C56)	7 415
		Naissuguelundid (muu) (C57)	316
		Platsenta (C58)	73
Meessuguelundid (C60–C63)	22 727	Suguti (C60)	371
		Eesnääre (C61)	21 539
		Munand (C62)	768
		Meessuguelundid (muu) (C63)	49
Kuseelundid (C64–C68)	17 872	Neer, neeruvaagen (C64–C65)	9 855
		Kusejuha (C66)	191
		Kusepõis (C67)	7 743
		Kuseelundid (muu) (C68)	83
Silm (C69)	718		
Peaaju ja kesknärvisüsteem (C70–C72)	4 085	Ajukelmed (C70)	97
		Peaaju (C71)	3 846
		Kesknärvisüsteem (muu) (C72)	142
Kilpnääre (C73)	2 629		
Sisesekreetsiooninäärmed (muu) (C74–C75)	321		
Paige teadmata või täpsustamata (C76–C80)	4 410		
Hodgkini tõbi (C81)	1 795		
Mitte-Hodgkini lümfoom (C82– C85/96)	5 103		
Immunoproliferatiivhaigused (C88)	73		
Hulgimüeloom (C90)	2 201		
Leukeemia (C91–C95)	7 912	Lümfoidleukeemia (C91)	4 485
		Müeloidleukeemia (C92)	2 535
		Leukeemia (muu) (C93–C95)	892
Lümfoid- ja vereloomekoe ning nendesarnaste kudede muud ning täpsustamata pahaloomulised kasvajaite (C96)	0		

## Lisa 2. Andmete ettevalmistamine (R)

```
library(readxl)

# Andmete sisselugemine
andmed <- read_excel("C:/Users/viktoria/Documents/Ülikool/Magister/Magistritöö/Viktoria Kirpu
magistritöök.xls")
names(andmed) <- c("Jrk", "SuguID", "Vanus", "Aasta", "Kuu", "Dgn", "MKkood", "MKnimi")

## Tunnuste formaadi muutmine
andmed$SuguID<- as.factor(andmed$SuguID) # factor
andmed$Vanus <- as.numeric(andmed$Vanus) # integer
andmed$Aasta <- as.numeric(andmed$Aasta) # integer
andmed$Kuu <- as.numeric(andmed$Kuu) # integer
andmed$Dgn <- as.character(andmed$Dgn) # character
andmed$MKkood <- as.factor(andmed$MKkood) # factor
andmed$MKnimi <- as.factor(andmed$MKnimi) # factor
andmed$Jrk <- as.factor(andmed$Jrk) # factor

# Tunnuste kirjeldused ja esmane ülevaade
summary(andmed)

## Sugu
andmed$Sugu <- "Mees"
andmed$Sugu[andmed$SuguID == 2] <- "Naine"
andmed$Sugu<-as.factor(andmed$Sugu)
summary(andmed$Sugu)

## Vanus
hist(andmed$Vanus, breaks=103)#; table(andmed$Vanus)

##Diagnos
andmed$Dgn2 <- as.integer(substr(andmed$Dgn, 2, 3))

### Grupeering 1
andmed$DgnGR1 <- "Muu"
andmed$DgnGR1[andmed$Dgn2>=0 & andmed$Dgn2<=14] <- "C00-C14"
andmed$DgnGR1[andmed$Dgn2>=15 & andmed$Dgn2<=26] <- "C15-C26"
andmed$DgnGR1[andmed$Dgn2>=30 & andmed$Dgn2<=39] <- "C30-C39"
andmed$DgnGR1[andmed$Dgn2>=40 & andmed$Dgn2<=41] <- "C40-C41"
andmed$DgnGR1[andmed$Dgn2==43] <- "C43"
andmed$DgnGR1[andmed$Dgn2==44] <- "C44"
andmed$DgnGR1[andmed$Dgn2>=45 & andmed$Dgn2<=49] <- "C45-C49"
andmed$DgnGR1[andmed$Dgn2==50] <- "C50"
andmed$DgnGR1[andmed$Dgn2>=51 & andmed$Dgn2<=58] <- "C51-C58"
andmed$DgnGR1[andmed$Dgn2>=60 & andmed$Dgn2<=63] <- "C60-C63"
andmed$DgnGR1[andmed$Dgn2>=64 & andmed$Dgn2<=68] <- "C64-C68"
andmed$DgnGR1[andmed$Dgn2==69] <- "C69"
andmed$DgnGR1[andmed$Dgn2>=70 & andmed$Dgn2<=72] <- "C70-C72"
andmed$DgnGR1[andmed$Dgn2==73] <- "C73"
andmed$DgnGR1[andmed$Dgn2>=74 & andmed$Dgn2<=75] <- "C74-C75"
andmed$DgnGR1[andmed$Dgn2>=76 & andmed$Dgn2<=80] <- "C76-C80"
andmed$DgnGR1[andmed$Dgn2==81] <- "C81"
andmed$DgnGR1[(andmed$Dgn2>=82 & andmed$Dgn2<=85) | andmed$Dgn2==96] <- "C82-C85/96"
andmed$DgnGR1[andmed$Dgn2==88] <- "C88"
andmed$DgnGR1[andmed$Dgn2==90] <- "C90"
andmed$DgnGR1[andmed$Dgn2>=91 & andmed$Dgn2<=95] <- "C91-C95"
andmed$DgnGR1[andmed$Dgn2==97] <- "C97"
table(andmed$DgnGR1)
andmed$DgnGR1nimi <- "Muu"
andmed$DgnGR1nimi[andmed$Dgn2>=0 & andmed$Dgn2<=14] <- "Huu, suuõõs, neel"
andmed$DgnGR1nimi[andmed$Dgn2>=15 & andmed$Dgn2<=26] <- "Seedeelundid"
andmed$DgnGR1nimi[andmed$Dgn2>=30 & andmed$Dgn2<=39] <- "Hingamis-, rindkeresised elundid"
andmed$DgnGR1nimi[andmed$Dgn2>=40 & andmed$Dgn2<=41] <- "Luu, liigeskõhr"
andmed$DgnGR1nimi[andmed$Dgn2==43] <- "Nahk - melanoom"
andmed$DgnGR1nimi[andmed$Dgn2==44] <- "Nahk (muu)"
andmed$DgnGR1nimi[andmed$Dgn2>=45 & andmed$Dgn2<=49] <- "Mesoteelkude ja pehmed koed"
andmed$DgnGR1nimi[andmed$Dgn2==50] <- "Rind"
andmed$DgnGR1nimi[andmed$Dgn2>=51 & andmed$Dgn2<=58] <- "Naissuguelundid"
andmed$DgnGR1nimi[andmed$Dgn2>=60 & andmed$Dgn2<=63] <- "Meessuguelundid"
andmed$DgnGR1nimi[andmed$Dgn2>=64 & andmed$Dgn2<=68] <- "kuseelundid"
andmed$DgnGR1nimi[andmed$Dgn2==69] <- "silma"
andmed$DgnGR1nimi[andmed$Dgn2>=70 & andmed$Dgn2<=72] <- "Peaaju ja kesknärvisüsteem"
andmed$DgnGR1nimi[andmed$Dgn2==73] <- "Kilpnääre"
andmed$DgnGR1nimi[andmed$Dgn2>=74 & andmed$Dgn2<=75] <- "Sisesekretsiooninäärmed (muu)"
andmed$DgnGR1nimi[andmed$Dgn2>=76 & andmed$Dgn2<=80] <- "Paige teadmata või täpsustamata"
andmed$DgnGR1nimi[andmed$Dgn2==81] <- "Hodgkini tõbi"
andmed$DgnGR1nimi[(andmed$Dgn2>=82 & andmed$Dgn2<=85) | andmed$Dgn2==96] <- "Mitte-Hodgkini
lümfoom"
andmed$DgnGR1nimi[andmed$Dgn2==88] <- "Immunoproliferatiivhaigused"
andmed$DgnGR1nimi[andmed$Dgn2==90] <- "Hulgimüeloom"
andmed$DgnGR1nimi[andmed$Dgn2>=91 & andmed$Dgn2<=95] <- "Leukeemia"
andmed$DgnGR1nimi[andmed$Dgn2==97] <- "Sõltumatud hulgipaikmed"
table(andmed$DgnGR1nimi)
```

```

### Grupeering 2
andmed$DgnGR2 <- "Muu"
andmed$DgnGR2[andmed$Dgn2==0] <- "C00"
andmed$DgnGR2[andmed$Dgn2>=1 & andmed$Dgn2<=2] <- "C01-C02"
andmed$DgnGR2[andmed$Dgn2>=3 & andmed$Dgn2<=6] <- "C03-C06"
andmed$DgnGR2[andmed$Dgn2>=7 & andmed$Dgn2<=8] <- "C07-C08"
andmed$DgnGR2[andmed$Dgn2>=9 & andmed$Dgn2<=10] <- "C09-C10"
andmed$DgnGR2[andmed$Dgn2==11] <- "C11"
andmed$DgnGR2[andmed$Dgn2>=12 & andmed$Dgn2<=13] <- "C12-C13"
andmed$DgnGR2[andmed$Dgn2==14] <- "C14"
andmed$DgnGR2[andmed$Dgn2==15] <- "C15"
andmed$DgnGR2[andmed$Dgn2==16] <- "C16"
andmed$DgnGR2[andmed$Dgn2==17] <- "C17"
andmed$DgnGR2[andmed$Dgn2==18] <- "C18"
andmed$DgnGR2[andmed$Dgn2>=19 & andmed$Dgn2<=21] <- "C19-C21"
andmed$DgnGR2[andmed$Dgn2==22] <- "C22"
andmed$DgnGR2[andmed$Dgn2>=23 & andmed$Dgn2<=24] <- "C23-C24"
andmed$DgnGR2[andmed$Dgn2==25] <- "C25"
andmed$DgnGR2[andmed$Dgn2==26] <- "C26"
andmed$DgnGR2[andmed$Dgn2>=30 & andmed$Dgn2<=31] <- "C30-C31"
andmed$DgnGR2[andmed$Dgn2==32] <- "C32"
andmed$DgnGR2[andmed$Dgn2>=33 & andmed$Dgn2<=34] <- "C33-C34"
andmed$DgnGR2[andmed$Dgn2>=37 & andmed$Dgn2<=38] <- "C37-C38"
andmed$DgnGR2[andmed$Dgn2==39] <- "C39"
andmed$DgnGR2[andmed$Dgn2>=40 & andmed$Dgn2<=41] <- "C40-C41"
andmed$DgnGR2[andmed$Dgn2==43] <- "C43"
andmed$DgnGR2[andmed$Dgn2==44] <- "C44"
andmed$DgnGR2[andmed$Dgn2>=45 & andmed$Dgn2<=49] <- "C45-C49"
andmed$DgnGR2[andmed$Dgn2==50] <- "C50"
andmed$DgnGR2[andmed$Dgn2>=51 & andmed$Dgn2<=52] <- "C51-C52"
andmed$DgnGR2[andmed$Dgn2==53] <- "C53"
andmed$DgnGR2[andmed$Dgn2==54] <- "C54"
andmed$DgnGR2[andmed$Dgn2==55] <- "C55"
andmed$DgnGR2[andmed$Dgn2==56] <- "C56"
andmed$DgnGR2[andmed$Dgn2==57] <- "C57"
andmed$DgnGR2[andmed$Dgn2==58] <- "C58"
andmed$DgnGR2[andmed$Dgn2==60] <- "C60"
andmed$DgnGR2[andmed$Dgn2==61] <- "C61"
andmed$DgnGR2[andmed$Dgn2==62] <- "C62"
andmed$DgnGR2[andmed$Dgn2==63] <- "C63"
andmed$DgnGR2[andmed$Dgn2>=64 & andmed$Dgn2<=65] <- "C64-C65"
andmed$DgnGR2[andmed$Dgn2==66] <- "C66"
andmed$DgnGR2[andmed$Dgn2==67] <- "C67"
andmed$DgnGR2[andmed$Dgn2==68] <- "C68"
andmed$DgnGR2[andmed$Dgn2==69] <- "C69"
andmed$DgnGR2[andmed$Dgn2==70] <- "C70"
andmed$DgnGR2[andmed$Dgn2==71] <- "C71"
andmed$DgnGR2[andmed$Dgn2==72] <- "C72"
andmed$DgnGR2[andmed$Dgn2==73] <- "C73"
andmed$DgnGR2[andmed$Dgn2>=74 & andmed$Dgn2<=75] <- "C74-C75"
andmed$DgnGR2[andmed$Dgn2>=76 & andmed$Dgn2<=80] <- "C76-C80"
andmed$DgnGR2[andmed$Dgn2==81] <- "C81"
andmed$DgnGR2[(andmed$Dgn2>=82 & andmed$Dgn2<=85) | andmed$Dgn2==96] <- "C82-C85/96"
andmed$DgnGR2[andmed$Dgn2==88] <- "C88"
andmed$DgnGR2[andmed$Dgn2==90] <- "C90"
andmed$DgnGR2[andmed$Dgn2==91] <- "C91"
andmed$DgnGR2[andmed$Dgn2==92] <- "C92"
andmed$DgnGR2[andmed$Dgn2>=93 & andmed$Dgn2<=95] <- "C93-C95"
andmed$DgnGR2[andmed$Dgn2==97] <- "C97"
table(andmed$DgnGR2)
andmed$DgnGR2nimi <- "Muu"
andmed$DgnGR2nimi[andmed$Dgn2==0] <- "Huu1"
andmed$DgnGR2nimi[andmed$Dgn2>=1 & andmed$Dgn2<=2] <- "Keel"
andmed$DgnGR2nimi[andmed$Dgn2>=3 & andmed$Dgn2<=6] <- "Ige, suupõhi jm"
andmed$DgnGR2nimi[andmed$Dgn2>=7 & andmed$Dgn2<=8] <- "Suured sõljenäärmed"
andmed$DgnGR2nimi[andmed$Dgn2>=9 & andmed$Dgn2<=10] <- "Mandel, suuneel"
andmed$DgnGR2nimi[andmed$Dgn2==11] <- "Ninaneel"
andmed$DgnGR2nimi[andmed$Dgn2>=12 & andmed$Dgn2<=13] <- "Pirnsopis, alaneel"
andmed$DgnGR2nimi[andmed$Dgn2==14] <- "Huu1, suuõõs, neel (muu)"
andmed$DgnGR2nimi[andmed$Dgn2==15] <- "Söogitoru"
andmed$DgnGR2nimi[andmed$Dgn2==16] <- "Magu"
andmed$DgnGR2nimi[andmed$Dgn2==17] <- "Peensool"
andmed$DgnGR2nimi[andmed$Dgn2==18] <- "Käärsool"
andmed$DgnGR2nimi[andmed$Dgn2>=19 & andmed$Dgn2<=21] <- "Pärasool jm"
andmed$DgnGR2nimi[andmed$Dgn2==22] <- "Maks jm"
andmed$DgnGR2nimi[andmed$Dgn2>=23 & andmed$Dgn2<=24] <- "Sapipõis jm"
andmed$DgnGR2nimi[andmed$Dgn2==25] <- "Kõhunääre"
andmed$DgnGR2nimi[andmed$Dgn2==26] <- "Seedelundid (muu)"
andmed$DgnGR2nimi[andmed$Dgn2>=30 & andmed$Dgn2<=31] <- "Ninaõõs, kõrv, ninakõrvalurked"
andmed$DgnGR2nimi[andmed$Dgn2==32] <- "Kõri"
andmed$DgnGR2nimi[andmed$Dgn2>=33 & andmed$Dgn2<=34] <- "Hingetoru, bronh, kops"
andmed$DgnGR2nimi[andmed$Dgn2>=37 & andmed$Dgn2<=38] <- "Tüümus, süda, keskseinand, pleura"
andmed$DgnGR2nimi[andmed$Dgn2==39] <- "Hingamiselundid jm"
andmed$DgnGR2nimi[andmed$Dgn2>=40 & andmed$Dgn2<=41] <- "Luu, liigeskõhr"
andmed$DgnGR2nimi[andmed$Dgn2==43] <- "Nahk - melanoom"
andmed$DgnGR2nimi[andmed$Dgn2==44] <- "Nahk (muu)"
andmed$DgnGR2nimi[andmed$Dgn2>=45 & andmed$Dgn2<=49] <- "Mesoteelkude ja pehmed koed"

```



```

andmed$DgnGR2nimi [andmed$Dgn2==50] <- "Rind"
andmed$DgnGR2nimi [andmed$Dgn2>=51 & andmed$Dgn2<=52] <- "Häbe, tupp"
andmed$DgnGR2nimi [andmed$Dgn2==53] <- "Emakakael"
andmed$DgnGR2nimi [andmed$Dgn2==54] <- "Emakakeha"
andmed$DgnGR2nimi [andmed$Dgn2==55] <- "Emakas (osa täpsustamata)"
andmed$DgnGR2nimi [andmed$Dgn2==56] <- "Munasari"
andmed$DgnGR2nimi [andmed$Dgn2==57] <- "Naissuguelundid (muu)"
andmed$DgnGR2nimi [andmed$Dgn2==58] <- "Platsenta"
andmed$DgnGR2nimi [andmed$Dgn2==60] <- "Suguti"
andmed$DgnGR2nimi [andmed$Dgn2==61] <- "Eesnääre"
andmed$DgnGR2nimi [andmed$Dgn2==62] <- "Munand"
andmed$DgnGR2nimi [andmed$Dgn2==63] <- "Meessuguelundid (muu)"
andmed$DgnGR2nimi [andmed$Dgn2>=64 & andmed$Dgn2<=65] <- "Neer, neeruvaagen"
andmed$DgnGR2nimi [andmed$Dgn2==66] <- "Kusejuha"
andmed$DgnGR2nimi [andmed$Dgn2==67] <- "Kusepõis"
andmed$DgnGR2nimi [andmed$Dgn2==68] <- "Kuseelundid (muu)"
andmed$DgnGR2nimi [andmed$Dgn2==69] <- "Sil"
andmed$DgnGR2nimi [andmed$Dgn2==70] <- "Ajukeimed"
andmed$DgnGR2nimi [andmed$Dgn2==71] <- "Peaaju"
andmed$DgnGR2nimi [andmed$Dgn2==72] <- "Kesknärvisüsteem (muu)"
andmed$DgnGR2nimi [andmed$Dgn2==73] <- "Kilpnääre"
andmed$DgnGR2nimi [andmed$Dgn2>=74 & andmed$Dgn2<=75] <- "Sisesekretsiooninäärmad (muu)"
andmed$DgnGR2nimi [andmed$Dgn2>=76 & andmed$Dgn2<=80] <- "Paige teadmata või täpsustamata"
andmed$DgnGR2nimi [andmed$Dgn2==81] <- "Hodgkini tõbi"
andmed$DgnGR2nimi [(andmed$Dgn2>=82 & andmed$Dgn2<=85) | andmed$Dgn2==96] <- "Mitte-Hodgkini
lümfoom"
andmed$DgnGR2nimi [andmed$Dgn2==88] <- "Immunoproliferatiivhaigused"
andmed$DgnGR2nimi [andmed$Dgn2==90] <- "Hulgimüeloom"
andmed$DgnGR2nimi [andmed$Dgn2==91] <- "Lümfoidleukeemia"
andmed$DgnGR2nimi [andmed$Dgn2==92] <- "Müeloidleukeemia"
andmed$DgnGR2nimi [andmed$Dgn2>=93 & andmed$Dgn2<=95] <- "Leukeemia (muu)"
andmed$DgnGR2nimi [andmed$Dgn2==97] <- "Sõltumatud hulgipaikmed"
table(andmed$DgnGR2nimi)

# Agregeerime
ajutine <- andmed
ajutine$juhtumitearv <- 1
ajutine$DgnGRUPP1 <- paste(ajutine$DgnGR1nimi, "(", ajutine$DgnGR1, ")"), sep="")
ajutine$DgnGRUPP2 <- paste(ajutine$DgnGR2nimi, "(", ajutine$DgnGR2, ")"), sep="")
ajutine1<-aggregate(juhtumitearv ~ DgnGRUPP1+DgnGR1+DgnGRUPP2+DgnGR2, data=ajutine, FUN=sum)
ajutine2<-aggregate(juhtumitearv ~ DgnGRUPP1+DgnGR1, data=ajutine, FUN=sum)

andmed$Jrk <- as.factor(andmed$Jrk)
andmed$Sugu <- as.factor(andmed$Sugu)
andmed$Vanus <- as.numeric(andmed$Vanus)
andmed$Aasta <- as.numeric(andmed$Aasta)
andmed$Kuu <- NULL # ei ole hetkel tarvis
andmed$Dgn <- as.character(andmed$Dgn)
andmed$MKkood <- as.factor(andmed$MKkood)
andmed$MKnimi <- as.factor(andmed$MKnimi)
andmed$SuguID <- as.factor(andmed$SuguID)
andmed$DgnGR1 <- as.factor(andmed$DgnGR1)
andmed$DgnGR1nimi <- as.factor(andmed$DgnGR1nimi)
andmed$DgnGR2 <- as.factor(andmed$DgnGR2)
andmed$DgnGR2nimi <- as.factor(andmed$DgnGR2nimi)

### Sorteerime konkreetset ravijuhud
vajalikuddiagnoosid <- c("Eesnääre", "Emakakael")

#### Eraldame vajalikud andmeid
vajalikudandmed1 <- andmed[(andmed$DgnGR2nimi %in% vajalikuddiagnoosid), ]
vajalikudandmed1$DgnGRnimi <- vajalikudandmed1$DgnGR2nimi
vajalikudandmed1$DgnGR <- vajalikudandmed1$DgnGR2
vajalikudandmed1$DgnGR1 <- NULL
vajalikudandmed1$DgnGR2 <- NULL
vajalikudandmed1$DgnGR1nimi <- NULL
vajalikudandmed1$DgnGR2nimi <- NULL
vajalikudandmed2 <- andmed
vajalikudandmed2$DgnGRnimi <- "Kõik paikmed"
vajalikudandmed2$DgnGR <- "C.."
vajalikudandmed2$DgnGR1 <- NULL
vajalikudandmed2$DgnGR2 <- NULL
vajalikudandmed2$DgnGR1nimi <- NULL
vajalikudandmed2$DgnGR2nimi <- NULL
vajalikudandmed <- rbind(vajalikudandmed1, vajalikudandmed2)
vajalikudandmed$Jrk <- NULL
vajalikudandmed$MKkood <- NULL
vajalikudandmed$MKnimi <- NULL

#-----
## Statistikaamet

### Impordime andmeid
rahvaarvud <-
read.table("C:/users/viktoria/Documents/ülikool/Magister/Magistritöö/rahvaarv.csv", header =
TRUE, sep=",")

```

```

### Muudame andmeid
names(rahvaarvud) <- c("Aasta", "Sugu", "Vanus", "Rahvaarv")
rahvaarvud$Sugu <- as.character(rahvaarvud$Sugu)
rahvaarvud$Sugu[rahvaarvud$Sugu == "Mehed"] <- "Mees"
rahvaarvud$Sugu[rahvaarvud$Sugu == "Naised"] <- "Naine"
rahvaarvud$Sugu <- as.factor(rahvaarvud$Sugu)
rahvaarvud$Vanus <- as.character(rahvaarvud$Vanus)
rahvaarvud$Vanus[rahvaarvud$Vanus=="85 ja vanemad"] <- "85"
rahvaarvud$Vanus <- as.numeric(rahvaarvud$Vanus)
rahvaarv <- aggregate(Rahvaarv ~ Aasta+Sugu+Vanus, data=rahvaarvud, FUN=sum)
rm(rahvaarvud)

#-----

# Ühendame omavahel vajalikud andmed

## Agregeerime vajalikud andmed
vajalikudandmed$Vanus[vajalikudandmed$Vanus>85] <- 85
kokkupanemiseks <- vajalikudandmed
# names(kokkupanemiseks)
kokkupanemiseks$juhtumitearv <- 1
kokkupanemiseks <- aggregate(juhtumitearv ~ Aasta+Sugu+Vanus+DgnGR+DgnGRnimi,
data=kokkupanemiseks, FUN=sum)

## Paneme andmed kokku
(diagnoosid <- unique(kokkupanemiseks$DgnGR))
length(diagnoosid)
diagnoosinimi1 <- unique(kokkupanemiseks$DgnGRnimi[kokkupanemiseks$DgnGR == diagnoosid[1]])
diagnoos1 <- diagnoosid[1]
diagnoosinimi2 <- unique(kokkupanemiseks$DgnGRnimi[kokkupanemiseks$DgnGR == diagnoosid[2]])
diagnoos2 <- diagnoosid[2]
diagnoosinimi3 <- unique(kokkupanemiseks$DgnGRnimi[kokkupanemiseks$DgnGR == diagnoosid[3]])
diagnoos3 <- diagnoosid[3]
andmed1 <- kokkupanemiseks[kokkupanemiseks$DgnGR == diagnoos1, ]
andmed2 <- kokkupanemiseks[kokkupanemiseks$DgnGR == diagnoos2, ]
andmed3 <- kokkupanemiseks[kokkupanemiseks$DgnGR == diagnoos3, ]
andmed1 <- merge(andmed1, rahvaarv, by = c("Aasta", "Vanus", "Sugu"), all.x = TRUE,
all.y=TRUE)
andmed1$DgnGR[is.na(andmed1$DgnGR)] <- diagnoos1
andmed1$DgnGRnimi[is.na(andmed1$DgnGRnimi)] <- diagnoosinimi1
andmed1$juhtumitearv[is.na(andmed1$juhtumitearv)] <- 0
andmed2 <- merge(andmed2, rahvaarv, by = c("Aasta", "Vanus", "Sugu"), all.x = TRUE,
all.y=TRUE)
andmed2$DgnGR[is.na(andmed2$DgnGR)] <- diagnoos2
andmed2$DgnGRnimi[is.na(andmed2$DgnGRnimi)] <- diagnoosinimi2
andmed2$juhtumitearv[is.na(andmed2$juhtumitearv)] <- 0
andmed3 <- merge(andmed3, rahvaarv, by = c("Aasta", "Vanus", "Sugu"), all.x = TRUE,
all.y=TRUE)
andmed3$DgnGR[is.na(andmed3$DgnGR)] <- diagnoos3
andmed3$DgnGRnimi[is.na(andmed3$DgnGRnimi)] <- diagnoosinimi3
andmed3$juhtumitearv[is.na(andmed3$juhtumitearv)] <- 0
koosandmed <- rbind(andmed1, andmed2, andmed3)
koosandmed$DgnGRnimi <- as.character(koosandmed$DgnGRnimi)
koosandmed$DgnGRnimi <- as.factor(koosandmed$DgnGRnimi)

# Salvestame andmed analüüsimiseks eraldi faili:
write.csv(koosandmed, "C:\\Users\\viktoriak\\Documents\\Ülikool\\Magister\\Magistritöö\\andmed_
rahvastikuga.csv", row.names = FALSE)

```

## Lisa 3. Andmete kirjeldav analüüs (R)

```

library(ggplot2)
library(dplyr)
options(scipen = 999)
library(ggpubr)

# Andmete sisselugemine
koosandmed
read.table("C:/Users/viktoriak/Documents/Ülikool/Magister/Magistritöö/andmed_rahvastikuga.csv"
, header = TRUE, sep=",")

# Teeme diagrammid

## Uurime, mis diagnoosidele me jooniseid teeme:
(diagnoosid <- unique(koosandmed$DgnGRnimi))

andmestik1 <- koosandmed[koosandmed$DgnGRnimi == diagnoosid[1], ]
andmestik1 <- aggregate(cbind(juhtumitearv) ~ Aasta+DgnGRnimi, data=andmestik1, FUN=sum)
andmestik2 <- koosandmed[koosandmed$DgnGRnimi == diagnoosid[2], ]

```

```

andmestik2 <- aggregate(cbind(juhtumitearv) ~ Aasta+DgnGRnimi, data=andmestik2, FUN=sum)
andmestik3 <- koosandmed[koosandmed$DgnGRnimi == diagnoosid[3], ]
andmestik3 <- aggregate(cbind(juhtumitearv) ~ Aasta+DgnGRnimi, data=andmestik3, FUN=sum)
andmestikAasta <- rbind(andmestik1, andmestik2, andmestik3)
andmestikAasta$juhtumid <- log(andmestikAasta$juhtumitearv)
exp(andmestikAasta$juhtumid)
andmestikAasta$juhtumid[andmestikAasta$juhtumid==-Inf]<- -0.1
andmestik1 <- koosandmed[koosandmed$DgnGRnimi == diagnoosid[1], ]
andmestik1 <- aggregate(cbind(juhtumitearv) ~ Vanus+DgnGRnimi, data=andmestik1, FUN=sum)
andmestik2 <- koosandmed[koosandmed$DgnGRnimi == diagnoosid[2], ]
andmestik2 <- aggregate(cbind(juhtumitearv) ~ Vanus+DgnGRnimi, data=andmestik2, FUN=sum)
andmestik3 <- koosandmed[koosandmed$DgnGRnimi == diagnoosid[3], ]
andmestik3 <- aggregate(cbind(juhtumitearv) ~ Vanus+DgnGRnimi, data=andmestik3, FUN=sum)
andmestikvanus <- rbind(andmestik1, andmestik2, andmestik3)
andmestikvanus$juhtumid <- log(andmestikvanus$juhtumitearv)
andmestikvanus$juhtumid[andmestikvanus$juhtumid==-Inf] <- -0.1

andmestikAasta$DgnGRnimi <- factor(andmestikAasta$DgnGRnimi,
                                  levels = c("Kõik paikmed", "Eesnääre", "Emakakael"),
                                  labels=c("Kõik paikmed", "Eesnääre", "Emakakael"))
ggplot(andmestikAasta, aes(x=as.numeric(Aasta), y=juhtumid, group=DgnGRnimi, colour=DgnGRnimi))
+
  geom_line(stat = "identity", size=1) + theme_light()+
  geom_hline(yintercept=0, color="grey60", linetype = "dashed")+
  xlab("Aasta") + ylab("vähijuhtude arv (log-skaala)")+
  scale_color_manual("vähi paige", values = c("Kõik paikmed"="#7AB800",
                                             "Eesnääre"="#4B92DB",
                                             "Emakakael"="#FF7900"))+
  scale_x_continuous(breaks = seq(1970, 2010, by=10),
                    labels = as.character(seq(1970, 2010, by=10)))+
  scale_y_continuous(breaks = c(0, log(c(10, 100, 1000, 10000))),
                    limits=c(0, log(10000)),
                    labels= as.character(c(0, 10, 100, 1000, 10000)))+
  theme(panel.grid.minor = element_blank(),
        text=element_text(family="serif", size=16),
        legend.text=element_text(size=12))

andmestikvanus$DgnGRnimi <- factor(andmestikvanus$DgnGRnimi,
                                  levels = c("Kõik paikmed", "Eesnääre", "Emakakael"),
                                  labels=c("Kõik paikmed", "Eesnääre", "Emakakael"))
ggplot(andmestikvanus, aes(x=as.numeric(Vanus), y=juhtumid, group=DgnGRnimi, colour=DgnGRnimi))
+
  geom_vline(mapping=NULL, xintercept=seq(0, 85, by=5), colour='grey90')+
  geom_line(stat = "identity", size=1) + theme_light()+
  geom_hline(yintercept=0, color="grey60", linetype = "dashed")+
  xlab("Vanus") + ylab("vähijuhtude arv (log-skaala)")+
  scale_color_manual("vähi paige", values = c("Kõik paikmed"="#7AB800",
                                             "Eesnääre"="#4B92DB",
                                             "Emakakael"="#FF7900"))+
  scale_x_continuous(breaks = seq(0, 85, by=5),
                    labels = c(as.character(seq(0, 80, by=5)), "85"))+
  scale_y_continuous(breaks = c(0, log(c(10, 100, 1000, 10000))),
                    labels= as.character(c(0, 10, 100, 1000, 10000)))+
  theme(panel.grid.minor = element_blank(),
        text=element_text(family="serif", size=16),
        legend.text=element_text(size=12))

diagnoos <- "Kõik paikmed"
#diagnoos <- "Eesnääre"
#diagnoos <- "Emakakael"
andmed <- koosandmed[koosandmed$DgnGRnimi == diagnoos, ]
if (diagnoos == "Emakakael" | diagnoos == "Eesnääre"){
  andmedAasta <- aggregate(cbind(juhtumitearv, Rahvaarv) ~ Aasta, data=andmed, FUN=sum)
  andmedVanus <- aggregate(cbind(juhtumitearv, Rahvaarv) ~ Vanus, data=andmed, FUN=sum)
  andmedAasta60 <- aggregate(cbind(juhtumitearv, Rahvaarv) ~ Aasta,
                           data=andmed[andmed$Vanus==60, ], FUN=sum)
  andmedVanus2000 <- aggregate(cbind(juhtumitearv, Rahvaarv) ~ Vanus,
                             data=andmed[andmed$Aasta==2000, ], FUN=sum)}
if (diagnoos == "Kõik paikmed"){
  andmedAasta <- aggregate(cbind(juhtumitearv, Rahvaarv) ~ Aasta+Sugu, data=andmed, FUN=sum)
  andmedVanus <- aggregate(cbind(juhtumitearv, Rahvaarv) ~ Vanus+Sugu, data=andmed, FUN=sum)
  andmedAasta60 <- aggregate(cbind(juhtumitearv, Rahvaarv) ~ Aasta+Sugu,
                             data=andmed[andmed$Vanus==60, ], FUN=sum)
  andmedVanus2000 <- aggregate(cbind(juhtumitearv, Rahvaarv) ~ Vanus+Sugu,
                              data=andmed[andmed$Aasta==2000, ], FUN=sum)}

for (i in 1:nrow(andmedVanus)){
  test<-binom.test(andmedVanus$juhtumitearv[i], andmedVanus$Rahvaarv[i])
  andmedVanus$alumine[i] <- test$conf.int[1]
  andmedVanus$ulemine[i] <- test$conf.int[2]}
for (i in 1:nrow(andmedAasta)){
  test<-binom.test(andmedAasta$juhtumitearv[i], andmedAasta$Rahvaarv[i])
  andmedAasta$alumine[i] <- test$conf.int[1]
  andmedAasta$ulemine[i] <- test$conf.int[2]}

if (diagnoos == "Kõik paikmed"){
  p1 <- ggplot(andmedVanus, aes(x = Vanus, y = juhtumitearv/Rahvaarv, fill = Sugu)) +

```

```

geom_bar(data = subset(andmedVanus, Sugu == "Mees"), stat = "identity", width = 1, alpha=0.7)
+
geom_bar(data = subset(andmedVanus, Sugu == "Naine"), stat = "identity", width = 1,
alpha=0.7) +
scale_x_continuous(breaks = seq(0, 85, by=5),
labels = c(as.character(seq(0, 80, by=5)), "85+"))+
theme(axis.ticks = element_blank()) +
xlab("Vanus") +
scale_fill_manual("Sugu", values = c("Mees"="#4B92DB", "Naine"="#FF7900"))+
theme_gray(base_size = 12)+theme_light()+
theme(panel.grid.minor = element_blank(),
axis.text.y=element_text(size=11),
axis.title.y=element_blank(),
text=element_text(family="serif", size=16),
legend.text=element_text(size=12))+
theme(axis.title.x=element_blank(),
axis.text.x=element_blank(),
axis.ticks.x=element_blank())
p2 <- ggplot(andmedVanus, aes(x = Vanus, y = juhtumitearv/Rahvaarv, color = Sugu)) +
geom_errorbar(aes(ymin = andmedVanus$alumine,ymax = andmedVanus$ulemine), width = 0.5,size
= 0.5, color = "grey")+
geom_point(data = subset(andmedVanus, Sugu == "Mees"), stat = "identity", width = 1, alpha=1)
+
geom_point(data = subset(andmedVanus, Sugu == "Naine"), stat = "identity", width = 1,
alpha=1) +
scale_x_continuous(breaks = seq(0, 85, by=5),
labels = c(as.character(seq(0, 80, by=5)), "85+"))+
theme(axis.ticks = element_blank()) +
xlab("Vanus") +ylim(0,max(andmedVanus$ulemine))+
theme(axis.text.y = element_text(size = 10))+
scale_color_manual("Sugu", values = c("Mees"="#4B92DB", "Naine"="#FF7900"))+
theme_gray(base_size = 12)+theme_light()+
theme(panel.grid.minor = element_blank(),
axis.text.x=element_text(size=11),
axis.text.y=element_text(size=11),
axis.title.y=element_blank(),
text=element_text(family="serif", size=16),
legend.text=element_text(size=12))
annotate_figure(ggarrange(p1,p2,ncol=1, nrow=2, heights=c(0.45,0.55),
common.legend = FALSE, legend="right"),
left = text_grob("Risk", rot = 90, size=16, family="serif"))
p1 <- ggplot(andmedAasta, aes(x = Aasta, y = juhtumitearv/Rahvaarv, fill = Sugu)) +
geom_bar(data = subset(andmedAasta, Sugu == "Mees"),
stat = "identity", width = 1, alpha=0.7) +
geom_bar(data = subset(andmedAasta, Sugu == "Naine"),
stat = "identity", width = 1, alpha=0.7) +
scale_x_continuous(breaks = seq(1970,2010, by=10),
labels = as.character(seq(1970,2010, by=10))) +
theme(axis.ticks = element_blank()) +
xlab("Aasta") +theme_light()+
scale_fill_manual("Sugu", values = c("Mees"="#4B92DB", "Naine"="#FF7900"))+
theme(panel.grid.minor = element_blank(),
text=element_text(family="serif", size=16),
legend.text=element_text(size=12))+
theme(axis.title.x=element_blank(),
axis.title.y=element_blank(),
axis.text.y=element_text(size=11),
axis.text.x=element_blank())
p2 <- ggplot(andmedAasta, aes(x = Aasta, y = juhtumitearv/Rahvaarv, color = Sugu)) +
geom_errorbar(aes(ymin = andmedAasta$alumine,ymax = andmedAasta$ulemine), width = 0.5,size
= 0.5, color = "grey")+
geom_point(data = subset(andmedAasta, Sugu == "Mees"), stat = "identity", width = 1, alpha=1)
+
geom_point(data = subset(andmedAasta, Sugu == "Naine"), stat = "identity", width = 1,
alpha=1) +
theme(axis.ticks = element_blank()) + theme_light()+
xlab("Aasta") + ylim(0,max(andmedAasta$ulemine))+
scale_color_manual("Sugu", values = c("Mees"="#4B92DB", "Naine"="#FF7900"))+
theme(panel.grid.minor = element_blank(),
axis.text.x=element_text(size=11),
axis.text.y=element_text(size=11),
axis.title.y=element_blank(),
text=element_text(family="serif", size=16),
legend.text=element_text(size=12))
annotate_figure(ggarrange(p1,p2,ncol=1, nrow=2, heights=c(0.45,0.55),
common.legend = FALSE, legend="right"),
left = text_grob("Risk", rot = 90, size=16, family="serif"))
p1 <- ggplot(andmedVanus2000, aes(x=as.numeric(Vanus), y=juhtumitearv, fill=Sugu)) +
geom_vline(mapping=NULL, xintercept=seq(0, 85, by=5),colour='grey90')+
geom_bar(data = subset(andmedVanus2000, Sugu == "Mees"), stat = "identity", width = 1,
alpha=0.7) +
geom_bar(data = subset(andmedVanus2000, Sugu == "Naine"), stat = "identity", width = 1,
alpha=0.7) +
scale_fill_manual("Sugu", values = c("Mees"="#4B92DB", "Naine"="#FF7900"))+
xlab("Vanus") + ylab("Vähijuhtude arv") + theme_light() +
theme(axis.ticks = element_blank()+
scale_x_continuous(breaks = seq(0, 85, by=5),

```

```

        labels = c(as.character(seq(0, 80, by=5)), "85+"))+
        scale_y_continuous(breaks = seq(0, 20*ceiling(max(andmedvanus2000$juhtumitearv)/20),
by=20),
        limits=c(0, max(andmedvanus2000$juhtumitearv))+
        theme(panel.grid.minor = element_blank(), axis.text.x=element_text(size=10),
axis.text.y=element_text(size=10), text=element_text(family="serif", size=16),
legend.text=element_text(size=12))
p2 <- ggplot(andmedAasta60, aes(x=as.numeric(Aasta), y=juhtumitearv, fill=Sugu)) +
geom_bar(data = subset(andmedAasta60, Sugu == "Mees"), stat = "identity", width = 1,
alpha=0.7) +
geom_bar(data = subset(andmedAasta60, Sugu == "Naine"), stat = "identity", width = 1,
alpha=0.7) +
scale_fill_manual("Sugu", values = c("Mees"="#4B92DB", "Naine"="#FF7900"))+
scale_x_continuous(breaks = seq(1970, 2010, by=10),
labels = as.character(seq(1970, 2010, by=10)))+
scale_y_continuous(breaks = seq(0, 10*ceiling(max(andmedAasta60$juhtumitearv)/10), by=10),
limits=c(0, max(andmedAasta60$juhtumitearv))+
xlab("Aasta") + theme_light() +
theme(panel.grid.minor = element_blank(), axis.title.y=element_blank(),
axis.ticks.y=element_blank(),axis.text.y=element_text(size=10),
axis.text.x=element_text(size=10), text=element_text(family="serif", size=16),
legend.text=element_text(size=12))
ggarrange(p1,p2,ncol=2, nrow=1, widths = c(1.5,1), common.legend = TRUE, legend="right")
}

if (diagnos == "Emakakael" | diagnos == "Eesnääre"){
if (diagnos == "Emakakael"){varvus <- "#FF7900"; vahe <- 1}
if (diagnos == "Eesnääre"){varvus <- "#4B92DB"; vahe <- 4}
ggplot(andmedVanus,
aes(x=as.numeric(Vanus), y=juhtumitearv/Rahvaarv)) +
geom_vline(mapping=NULL, xintercept=seq(0, 85, by=5),colour='grey90')+
geom_bar(stat = "identity", fill=varvus, alpha=0.7, width=1) +
xlab("Vanus") + ylab("Risk")+ theme_light()+
theme(axis.ticks = element_blank()+
scale_x_continuous(breaks = seq(0, 85, by=5),
labels = c(as.character(seq(0, 80, by=5)), "85+"))+
theme(panel.grid.minor = element_blank(),
axis.text.x=element_text(size=11),
axis.text.y=element_text(size=11),
text=element_text(family="serif", size=16),
legend.text=element_text(size=12))
ggplot(andmedAasta, aes(x=as.numeric(Aasta), y=juhtumitearv/Rahvaarv)) +
geom_bar(stat = "identity", fill=varvus, width=1, alpha=0.7) +
xlab("Aasta") + ylab("Risk")+
theme(axis.ticks = element_blank()+
scale_x_continuous(breaks = seq(1970, 2010, by=10),
labels = as.character(seq(1970, 2010, by=10)))+
theme_gray(base_size = 12)+theme_light()+
theme(panel.grid.minor = element_blank(),
axis.text.x=element_text(size=11),
axis.text.y=element_text(size=11),
text=element_text(family="serif", size=16),
legend.text=element_text(size=12))
p1 <- ggplot(andmedvanus2000, aes(x=as.numeric(Vanus), y=juhtumitearv)) + theme_light() +
geom_vline(mapping=NULL, xintercept=seq(0, 85, by=5),colour='grey90')+
geom_bar(stat = "identity", fill=varvus, alpha=0.7, width=1) +
xlab("Vanus") + ylab("Vähijuhtude arv") +
theme(axis.ticks = element_blank()+
scale_x_continuous(breaks = seq(0, 85, by=5),
labels = c(as.character(seq(0, 80, by=5)), "85+"))+
scale_y_continuous(breaks = seq(0, vahe*ceiling(max(andmedvanus2000$juhtumitearv)/vahe),
by=vahe),
limits=c(0, max(andmedvanus2000$juhtumitearv))+
theme(text=element_text(family="serif")) +
theme(panel.grid.minor = element_blank(), axis.text.x=element_text(size=10),
axis.text.y=element_text(size=10), text=element_text(family="serif", size=16),
legend.text=element_text(size=12))
p2 <- ggplot(andmedAasta60, aes(x=as.numeric(Aasta), y=juhtumitearv)) +
geom_bar(stat = "identity", fill=varvus, width=1, alpha=0.7) +
scale_x_continuous(breaks = seq(1970, 2010, by=10),
labels = as.character(seq(1970, 2010, by=10)))+
scale_y_continuous(breaks = seq(0, vahe*ceiling(max(andmedAasta60$juhtumitearv)/vahe),
by=vahe),
limits=c(0, max(andmedAasta60$juhtumitearv))+
xlab("Aasta") + theme_light() +
theme(panel.grid.minor = element_blank(), axis.title.y=element_blank(),
axis.ticks.y=element_blank(),axis.text.y=element_text(size=10),
axis.text.x=element_text(size=10), text=element_text(family="serif", size=16),
legend.text=element_text(size=12))
ggarrange(p1,p2,ncol=2, nrow=1, common.legend = TRUE, legend="right", widths = c(1.5,1))
}

```

## Lisa 4. Parima mudeli valimine (R)

```
library(splines)
options(scipen = 999)

# Andmete ettevalmistamine

koosandmed <-
read.table("C:/Users/viktoriak/Documents/Ülikool/Magister/Magistritöö/andmed_rahvastikuga.csv"
, header = TRUE, sep=",")
diagnoos <- "Kõik paikmed"
#diagnoos <- "Emakakael"
#diagnoos <- "Eesnääre"
data<-aggregate(cbind(juhtumitearv, Rahvaarv)~Vanus + Sugu + Aasta,
                data=koosandmed[(koosandmed$DgnGRnimi == diagnoos & koosandmed$Aasta != 2017),
], FUN=sum)
if (diagnoos == "Emakakael") {data <- data[data$Sugu == "Naine", ]}
if (diagnoos == "Eesnääre") {data <- data[data$Sugu == "Mees", ]}
data$Risk <- data$juhtumitearv/data$Rahvaarv
data <- data[order(data$Aasta, data$Sugu, data$Vanus), ]

# Parima splaini valik

## Vanus
splain1 <- glm(juhtumitearv~
              Sugu + # emakakaela ja eesnäärme puhul välja kommenteeritud
              bs(Vanus, df=0 + 3, degree=3) +
              bs(Vanus, df=0 + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
kommenteeritud
              bs(Aasta, df=3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
kommenteeritud
              bs(Aasta, df=3, degree=3),
              data = data, offset=log(Rahvaarv), family=poisson("log"))
splain2 <- glm(juhtumitearv~
              Sugu + # emakakaela ja eesnäärme puhul välja kommenteeritud
              bs(Vanus, df=1 + 3, degree=3) +
              bs(Vanus, df=1 + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
kommenteeritud
              bs(Aasta, df=3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
kommenteeritud
              bs(Aasta, df=3, degree=3),
              data = data, offset=log(Rahvaarv), family=poisson("log"))
splain3 <- glm(juhtumitearv~
              Sugu + # emakakaela ja eesnäärme puhul välja kommenteeritud
              bs(Vanus, df=2 + 3, degree=3) +
              bs(Vanus, df=2 + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
kommenteeritud
              bs(Aasta, df=3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
kommenteeritud
              bs(Aasta, df=3, degree=3),
              data = data, offset=log(Rahvaarv), family=poisson("log"))
muutus1 <- splain1$aic-splain2$aic
muutus2 <- splain1$aic-splain3$aic
splain <- splain1
n<-3
if (muutus1>0 | muutus2>0){
  splain <- splain2
  while (TRUE){
    splain1 <- splain2
    splain2 <- splain3
    splain3 <- glm(juhtumitearv~
                  Sugu + # emakakaela ja eesnäärme puhul välja kommenteeritud
                  bs(Vanus, df=n + 3, degree=3) +
                  bs(Vanus, df=n + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul
välja kommenteeritud
                  bs(Aasta, df=3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
kommenteeritud
                  bs(Aasta, df=3, degree=3),
                  data = data, offset=log(Rahvaarv), family=poisson("log"))
    muut1 <- splain1$aic-splain2$aic
    muut2 <- splain1$aic-splain3$aic
    if (muut1>0){
      splain <- splain2
      n <- n + 1
      next
    } else if (muut2>0){
      n <- n + 1
      next
    } else{
      n <- n-2
      break
    }
  }
} else {
  n <- 0
}
```

```

}
summary(splain)
n # katkemispunktide arv vanusele

## Aasta
splain1 <- glm(juhtumitearv~
  Sugu + # emakakaela ja eesnäärme puhul välja kommenteeritud
  bs(Vanus, df=n + 3, degree=3) +
  bs(Vanus, df=n + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
kommenteeritud
kommenteeritud
  bs(Aasta, df=0 + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
splain2 <- glm(juhtumitearv~
  data = data, offset=log(Rahvaarv), family=poisson("log"))
  Sugu + # emakakaela ja eesnäärme puhul välja kommenteeritud
  bs(Vanus, df=n + 3, degree=3) +
  bs(Vanus, df=n + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
kommenteeritud
kommenteeritud
  bs(Aasta, df=1 + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
splain3 <- glm(juhtumitearv~
  data = data, offset=log(Rahvaarv), family=poisson("log"))
  Sugu + # emakakaela ja eesnäärme puhul välja kommenteeritud
  bs(Vanus, df=n + 3, degree=3) +
  bs(Vanus, df=n + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
kommenteeritud
kommenteeritud
  bs(Aasta, df=2 + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul välja
  bs(Aasta, df=2 + 3, degree=3),
  data = data, offset=log(Rahvaarv), family=poisson("log"))
muutus1 <- splain1$aic-splain2$aic
muutus2 <- splain1$aic-splain3$aic
splain <- splain1
m<-3
if (muutus1>0 | muutus2>0){
  splain <- splain2
  while (TRUE){
    splain1 <- splain2
    splain2 <- splain3
    splain3 <- glm(juhtumitearv~
      Sugu + # emakakaela ja eesnäärme puhul välja kommenteeritud
      bs(Vanus, df=n + 3, degree=3) +
      bs(Vanus, df=n + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul
välja kommenteeritud
välja kommenteeritud
      bs(Aasta, df=m + 3, degree=3)*Sugu + # emakakaela ja eesnäärme puhul
      bs(Aasta, df=m + 3, degree=3),
      data = data, offset=log(Rahvaarv), family=poisson("log"))
    muut1 <- splain1$aic-splain2$aic
    muut2 <- splain1$aic-splain3$aic
    if (muut1>0){
      splain <- splain2
      m <- m + 1
      next
    } else if (muut2>0){
      m <- m + 1
      next
    } else{
      m <- m-2
      break
    }
  }
} else {
  m <- 0
}
summary(splain)
m # katkemispunktide arv Aastale

```

## Lisa 5. Mudeli konstrueerimine ja analüüs (R)

```

library(MASS)
library(gmodels)
library(ggplot2)

options(scipen = 999)
knitr::opts_chunk$set(fig.width=8, fig.height=6)

# MUUTUVAD PARAMEETRID

```

```

VanusProov <- 60
AastaProov <- 2000

# Andmete ettevalmistamine
koosandmed <-
read.table("C:/Users/viktoria/Documents/Ülikool/Magister/Magistritöö/andmed_rahvastikuga.csv",
, header = TRUE, sep=",")
diagnoos <- "Kõik paikmed"; n<-10; m<-9
#diagnoos <- "Emakakael"; n<-2; m<-3
#diagnoos <- "Eesnääre"; n<-1; m<-2
data<-aggregate(cbind(juhtumitearv, Rahvaarv)~Vanus + Sugu + Aasta,
                data=koosandmed[(koosandmed$DgnGRnimi == diagnoos & koosandmed$Aasta != 2017),
], FUN=sum)
if (diagnoos == "Emakakael") {data <- data[data$Sugu == "Naine", ]}
if (diagnoos == "Eesnääre") {data <- data[data$Sugu == "Mees", ]}
data$Risk <- data$juhtumitearv/data$Rahvaarv
data <- data[order(data$Aasta, data$Sugu, data$Vanus), ]

# Mudel
katkemispunktid1 <- c()
for (i in 1:n){
  katkemispunktid1 <- c(katkemispunktid1, as.numeric(quantile(data$Vanus, i/(1 + n))))
}
katkemispunktid2 <- c()
for (i in 1:m){
  katkemispunktid2 <- c(katkemispunktid2, as.numeric(quantile(data$Aasta, i/(1 + m))))
}
mudelistring <- paste("juhtumitearv~",
                    "Sugu + ", # emakakaela ja eesnäärme puhul välja kommenteeritud
                    "Vanus + I(Vanus**2) + I(Vanus**3)", sep="")

if (n!=0){
  for (i in 1:(n)){
    ,, mudelistring <- paste(mudelistring, paste("I((Vanus-", katkemispunktid1[i], ")**3*(Vanus-
    ,, katkemispunktid1[i], ">0))", sep=""), sep=" + ")
  }
}
# järgnevad 6 rida emakakaela ja eesnäärme puhul välja kommenteeritud
mudelistring <- paste(mudelistring, "Vanus*Sugu + I(Vanus**2)*Sugu + I(Vanus**3)*Sugu", sep="
+ ")
if (n!=0){
  for (i in 1:(n)){
    ,, mudelistring <- paste(mudelistring, paste("I((Vanus-", katkemispunktid1[i], ")**3*(Vanus-",
    ,, katkemispunktid1[i], ">0))*Sugu", sep=""), sep=" + ")
  }
}
mudelistring <- paste(mudelistring, "Aasta + I(Aasta**2) + I(Aasta**3)", sep=" + ")
if (m!=0){
  for (i in 1:(m)){
    ,, mudelistring <- paste(mudelistring, paste("I((Aasta-", katkemispunktid2[i], ")**3*(Aasta-
    ,, katkemispunktid2[i], ">0))", sep=""), sep=" + ")
  }
}
# järgnevad 6 rida emakakaela ja eesnäärme puhul välja kommenteeritud
mudelistring <- paste(mudelistring, "Aasta*Sugu + I(Aasta**2)*Sugu + I(Aasta**3)*Sugu", sep="
+ ")
if (m!=0){
  for (i in 1:(m)){
    ,, mudelistring <- paste(mudelistring, paste("I((Aasta-", katkemispunktid2[i], ")**3*(Aasta-",
    ,, katkemispunktid2[i], ">0))*Sugu", sep=""), sep="
+ ")
  }
}
mudel <- glm(mudelistring, data = data, offset=log(Rahvaarv), family=poisson("log"))
summary(mudel)

## Ülehajuvus?
(hajuvus <- sum(residuals(mudel,type="pearson")^2)/mudel$df.residual)
if (hajuvus > 2){
  mudel <- update(mudel, family=quasipoisson("log"))
  summary(mudel)
  sum(residuals(mudel,type="pearson")^2)/mudel$df.residual
  print("Kasutatud quasipoisson")
}

andmedAasta <- data[data$Vanus==VanusProov, ]
for (i in 1:nrow(andmedAasta)){
  andmedAasta$Risk[i] <- binom.test(andmedAasta$juhtumitearv[i],
andmedAasta$Rahvaarv[i])$estimate
}
andmedVanus <- data[data$Aasta== AastaProov, ]
for (i in 1:nrow(andmedVanus)){
  andmedVanus$Risk[i] <- binom.test(andmedVanus$juhtumitearv[i],
andmedVanus$Rahvaarv[i])$estimate
}

```



```

# Visualiseerimine
## VANUS
andmestik1 <- data.frame(Vanus=seq(0, 85, length.out =86), Rahvaarv = rep(10000, 86),
                        Sugu=rep("Mees", 86),
                        Aasta = rep(AastaProov, 86))
andmestik2 <- data.frame(Vanus=seq(0, 85, length.out =86), Rahvaarv = rep(10000, 86),
                        Sugu=rep("Naine", 86),
                        Aasta = rep(AastaProov, 86))
if (diagnoos == "Emakakael") {andmestik <- andmestik2
} else if(diagnoos == "Eesnääre") {andmestik <- andmestik1
} else {andmestik <- rbind(andmestik1, andmestik2)}

### Leia me suurused, mida paneme joonisele
#### Mudeli usaldusintervall (Joonis 1)
andmedVanus$predicted <- predict(mudel, andmestik, type="response")/10000

#### Punktide musaldusintervallid (Joonis 2)
andmedVanus$ulemine <- 0
andmedVanus$alumine <- 0
for (i in 1:nrow(andmedVanus)){
  test<-binom.test(andmedVanus$juhtumitearv[i], andmedVanus$Rahvaarv[i])
  andmedVanus$alumine[i] <- test$conf.int[1]
  andmedVanus$ulemine[i] <- test$conf.int[2]
}

#### Statistiliselt olulised muutused (Joonis 3)
##### Naine
if (diagnoos != "Eesnääre"){
  visualiseerimiseks1 <- data.frame(Vanus = seq(0, 85, length.out =86),
                                   Risk = predict(mudel, andmestik[andmestik$Sugu=="Naine", ],
                                                  type="response")/10000)
  juhtumitearvud <- data[data$Aasta==AastaProov & data$Sugu=="Naine", c(1,4)]
  algus <- 0
  lopp <- 0
  for(p in 1:nrow(juhtumitearvud)){
    if (juhtumitearvud$juhtumitearv[p]==0){
      lopp <- p
    }
    else{break}
  }
  visualiseerimiseks1$Sugu <- "Naine"
  visualiseerimiseks1$Varvid <- ""
  visualiseerimiseks1$Varvid[algus:lopp] <- "#003145"
  katkemispunktid <- c()
  for (i in 1:n){
    katkemispunktid <- c(katkemispunktid, as.numeric(quantile(data$Vanus, i/(1+n))))
  }
  for(i in 1:86){
    j <- i-1
    vektor <- c(0,
               0, # emakakaela ja eesnäärme puhul välja kommenteeritud
               1, 2*j, 3*j**2)
    if (n!=0){
      for(k in 1:n){
        vektor <- c(vektor, 3*(j-katkemispunktid[k])**2*I(j-katkemispunktid[k]>0))
      }
    }
    vektor <- c(vektor, rep(0, m+3))
    # järgmised 7 rida on emakakaela ja eesnäärme puhul välja kommenteeritud
    vektor <- c(vektor, 1, 2*j, 3*j**2)
    if (n!=0){
      for(k in 1:n){
        vektor <- c(vektor, 3*(j-katkemispunktid[k])**2*I(j-katkemispunktid[k]>0))
      }
    }
    vektor <- c(vektor, rep(0, m+3))
    ennustus <- estimable(mudel, vektor)
    if (ennustus$Estimate>0 & ennustus[1,5]< 0.05 & visualiseerimiseks1$Varvid[i] == ""){
      visualiseerimiseks1$Varvid[i]<-"#ff3424"
    } else if (ennustus$Estimate<0 & ennustus[1,5] < 0.05 & visualiseerimiseks1$Varvid[i] ==
    ""){
      visualiseerimiseks1$Varvid[i]<-"#92dd00"
    } else {visualiseerimiseks1$Varvid[i]<-"#003145"}
  }
}
##### Mees
if (diagnoos != "Emakakael"){
  visualiseerimiseks2 <- data.frame(Vanus = seq(0, 85, length.out =86),
                                   Risk = predict(mudel, andmestik[andmestik$Sugu=="Mees", ],
                                                  type="response")/10000)
  juhtumitearvud <- data[data$Aasta==AastaProov & data$Sugu=="Mees", c(1,4)]
  algus <- 0
  lopp <- 0

```

```

for(p in 1:nrow(juhtumitearvud)){
  if (juhtumitearvud$juhtumitearv[p]==0){
    lopp <- p
  }
  else{break}
}
visualiseerimiseks2$Sugu <- "Mees"
visualiseerimiseks2$Varvid <- ""
visualiseerimiseks2$Varvid[algus:lopp] <- "#003145"
katkemispunktid <- c()
for (i in 1:n){
  katkemispunktid <- c(katkemispunktid, as.numeric(quantile(data$Vanus, i/(1+n))))
}
for(i in 1:86){
  j <- i-1
  vektor <- c(0,
              0, # emakakaela ja eesnäärme puhul välja kommenteeritud
              1, 2*j, 3*j**2)

  if (n!=0){
    for(k in 1:n){
      vektor <- c(vektor, 3*(j-katkemispunktid[k])**2*I(j-katkemispunktid[k]>0))
    }
  }
  vektor <- c(vektor, rep(0, m+3))
  # järgmised 7 rida on emakakaela ja eesnäärme puhul välja kommenteeritud
  vektor <- c(vektor, 0,0,0)
  if (n!=0){
    for(k in 1:n){
      vektor <- c(vektor, 0)
    }
  }
  vektor <- c(vektor, rep(0, m+3))
  ennustus <- estimable(mudel, vektor)
  if (ennustus$Estimate>0 & ennustus[1,5]< 0.05 & visualiseerimiseks2$Varvid[i] == ""){
    visualiseerimiseks2$Varvid[i]<-"#ff3424"
  }else if (ennustus$Estimate<0 & ennustus[1,5] < 0.05 & visualiseerimiseks2$Varvid[i] ==
  ""){
    visualiseerimiseks2$Varvid[i]<-"#92dd00"
  }else {visualiseerimiseks2$Varvid[i]<-"#003145"}
}
}
if (diagnoos == "Emakakaela") {visualiseerimiseks <- visualiseerimiseks1
} else if(diagnoos == "Eesnäärme") {visualiseerimiseks <- visualiseerimiseks2
} else {visualiseerimiseks <- rbind(visualiseerimiseks1, visualiseerimiseks2)}

### Joonised

#### JOONIS 1
ggplot(andmedVanus, aes(x=Vanus, y=Risk)) +
  facet_grid(cols = vars(Sugu))+ # emakakaela ja eesnäärme puhul välja kommenteeritud
  geom_vline(mapping=NULL, xintercept=seq(0, 85, by=5),colour='grey90')+
  geom_point(shape=16, size=1.5)+ theme_light()+ ylim(0, max(andmedVanus$Risk))+
  geom_line(aes(x= Vanus, y = predicted), color = "#D52B1E", linetype = "solid", size=0.8)+
  scale_x_continuous(breaks = c(seq(0, 85, by=5)),
                    labels = c(as.character(seq(0, 80, by=5)), "85"))+
  theme(panel.grid.minor = element_blank(),
        axis.text.x=element_text(size=12),
        axis.text.y=element_text(size=12),
        text=element_text(family="serif", size=16))+
  theme(strip.text.x = element_text(size = 20),
        strip.text.y = element_text(size = 20))

#### JOONIS 2
ggplot(data=andmedVanus, aes(x=Vanus, y=Risk)) +
  geom_vline(mapping=NULL, xintercept=seq(0, 85, by=5),colour='grey90')+
  facet_grid(Sugu ~ ., scales="free")+ # emakakaela ja eesnäärme puhul välja kommenteeritud
  geom_errorbar(aes(ymin = andmedVanus$alumine,ymax = andmedVanus$ulemine),
               width = 0.5,size = 0.5, color = "grey") +
  geom_point(shape = 16, size = 2)+theme_light()+
  geom_line(aes(x=Vanus, y = predicted), color = "#D52B1E", linetype = "solid", size=1)+
  geom_hline(yintercept=0, color="grey60", linetype = "dashed")+
  scale_x_continuous(breaks = seq(0, 85, by=5),
                    labels = c(as.character(seq(0, 80, by=5)), "85"))+
  theme(panel.grid.minor = element_blank(),
        axis.text.x=element_text(size=12),
        axis.text.y=element_text(size=12))+
  theme(text=element_text(family="serif", size=16))+
  theme(strip.text.x = element_text(size = 20),
        strip.text.y = element_text(size = 20))

#### JOONIS 3
ggplot(visualiseerimiseks, aes(x=Vanus, y=Risk)) +
  facet_grid(rows = vars(Sugu))+ # emakakaela ja eesnäärme puhul välja kommenteeritud
  geom_vline(mapping=NULL, xintercept=seq(0, 85, by=5),colour='grey90')+
  geom_line() + theme_light()+ ylim(0, max(andmedVanus$Risk))+
  geom_point(shape=21, fill=visualiseerimiseks$Varvid, size=3)+
  geom_hline(yintercept=0, color="grey60", linetype = "dashed")+

```

```

scale_x_continuous(breaks = seq(0, 85, by=5),
                  labels = c(as.character(seq(0, 80, by=5)), "85+"))+
theme(panel.grid.minor = element_blank(),
      axis.text.x=element_text(size=12),
      axis.text.y=element_text(size=12),
      axis.ticks = element_blank()+
      theme(text=element_text(family="serif", size=16))+
      theme(strip.text.x = element_text(size = 20),
            strip.text.y = element_text(size = 20))

## AASTA

andmestik1 <- data.frame(Aasta=1968:2016, Rahvaarv = rep(10000, 49),
                        Sugu=rep("Mees", 49),
                        Vanus = rep(VanusProov, 49))
andmestik2 <- data.frame(Aasta=1968:2016, Rahvaarv = rep(10000, 49),
                        Sugu=rep("Naine", 49),
                        Vanus = rep(VanusProov, 49))
if (diagnoos == "Emakakael") {andmestik <- andmestik2
} else if(diagnoos == "Eesnääre") {andmestik <- andmestik1
} else {andmestik <- rbind(andmestik1, andmestik2)}

### Leiame suurused, mida paneme joonisele

#### Mudeli usaldusintervall (Joonis 1)
andmedAasta <- andmedAasta[order(andmedAasta$Sugu, andmedAasta$Aasta), ]
andmedAasta$predicted <- predict(mudel, andmestik, type="response")/10000

#### Punktide musaldusintervallid (Joonis 2)
andmedAasta$ulemine <- 0
andmedAasta$alumine <- 0
for (i in 1:nrow(andmedAasta)){
  test<-binom.test(andmedAasta$juhtumitearv[i], andmedAasta$Rahvaarv[i])
  andmedAasta$alumine[i] <- test$conf.int[1]
  andmedAasta$ulemine[i] <- test$conf.int[2]
}

#### Statistiliselt olulised muutused (Joonis 3)
##### Naine
if (diagnoos != "Eesnääre"){
  visualiseerimiseks1 <- data.frame(Aasta = 1968:2016,
                                   Risk = predict(mudel, andmestik[andmestik$Sugu=="Naine",
]), type="response")/10000)
  juhtumitearvud <- data[data$Vanus==VanusProov & data$Sugu=="Naine", c(1,4)]
  algus <- 0
  lopp <- 0
  for(p in 1:nrow(juhtumitearvud)){
    if (juhtumitearvud$juhtumitearv[p]==0){
      lopp <- p
    }
    else{break}
  }
  visualiseerimiseks1$Sugu <- "Naine"
  visualiseerimiseks1$Varvid <- ""
  visualiseerimiseks1$Varvid[algus:lopp] <- "#003145"
  katkemispunktid <- c()
  for (i in 1:m){
    katkemispunktid <- c(katkemispunktid, as.numeric(quantile(data$Aasta, i/(1+m))))
  }
  for(i in 1:49){
    j <- visualiseerimiseks1$Aasta[i]
    vektor <- c(0,
               0, # emakakaela ja eesnäärme puhul välja kommenteeritud
               rep(0, n+3), 1, 2*j, 3*j**2)

    if(m!=0){
      for(k in 1:m){
        vektor <- c(vektor, 3*(j-katkemispunktid[k])**2*I(j-katkemispunktid[k]>0))
      }
    }
    # järgmised 7 rida on emakakaela ja eesnäärme puhul välja kommenteeritud
    vektor <- c(vektor, rep(0, n+3))
    vektor <- c(vektor, 1, 2*j, 3*j**2)
    if(m!=0){
      for(k in 1:m){
        vektor <- c(vektor, 3*(j-katkemispunktid[k])**2*I(j-katkemispunktid[k]>0))
      }
    }
    ennustus <- estimable(mudel, vektor)
    if (ennustus$Estimate>0 & ennustus[1,5]< 0.05 & visualiseerimiseks1$Varvid[i] == ""){
      visualiseerimiseks1$Varvid[i]<-"#ff3424"
    } else if (ennustus$Estimate<0 & ennustus[1,5] < 0.05 & visualiseerimiseks1$Varvid[i] ==
    ""){
      visualiseerimiseks1$Varvid[i]<-"#92dd00"
    } else {
      visualiseerimiseks1$Varvid[i]<-"#003145"
    }
  }
}

```

```

}
#### Mees
if (diagnos != "Emakakael"){
  visualiseerimiseks2 <- data.frame(Aasta = 1968:2016,
                                   Risk = predict(mudel, andmestik[andmestik$Sugu=="Mees", ],
type="response")/10000)
  juhtumitearvud <- data[data$vanus==VanusProov & data$Sugu=="Mees", c(1,4)]
  algus <- 0
  lopp <- 0
  for(p in 1:nrow(juhtumitearvud)){
    if (juhtumitearvud$juhtumitearv[p]==0){
      lopp <- p
    }
    else{break}
  }
  visualiseerimiseks2$Sugu <- "Mees"
  visualiseerimiseks2$Varvid <- ""
  visualiseerimiseks2$Varvid[algus:lopp] <- "#003145"

  katkemispunktid <- c()
  for (i in 1:m){
    katkemispunktid <- c(katkemispunktid, as.numeric(quantile(data$Aasta, i/(1+m))))
  }
  for(i in 1:49){
    j <- visualiseerimiseks2$Aasta[i]
    vektor <- c(0,
                0, # emakakaela ja eesnäärme puhul välja kommenteeritud
                rep(0, n+3), 1, 2*j, 3*j**2)

    if(m!=0){
      for(k in 1:m){
        vektor <- c(vektor, 3*(j-katkemispunktid[k])**2*I(j-katkemispunktid[k]>0))
      }
    }
    # järgmised 7 rida on emakakaela ja eesnäärme puhul välja kommenteeritud
    vektor <- c(vektor, rep(0, n+3))
    vektor <- c(vektor, 0, 0, 0)
    if(m!=0){
      for(k in 1:m){
        vektor <- c(vektor, 0)
      }
    }
    ennustus <- estimable(mudel, vektor)
    if (ennustus$Estimate>0 & ennustus[1,5]< 0.05 & visualiseerimiseks2$Varvid[i] == ""){
      visualiseerimiseks2$Varvid[i]<-"#ff3424"
    } else if (ennustus$Estimate<0 & ennustus[1,5] < 0.05 & visualiseerimiseks2$Varvid[i] ==
""){
      visualiseerimiseks2$Varvid[i]<-"#92dd00"
    } else {
      visualiseerimiseks2$Varvid[i]<-"#003145"
    }
  }
}
if (diagnos == "Emakakael") {visualiseerimiseks <- visualiseerimiseks1
} else if (diagnos == "Eesnäär") {visualiseerimiseks <- visualiseerimiseks2
} else {visualiseerimiseks <- rbind(visualiseerimiseks2, visualiseerimiseks1)}

### Joonised

#### JOONIS 1
ggplot(andmedAasta, aes(x=Aasta, y=Risk)) +
  facet_grid(cols = vars(Sugu))+ # emakakaela ja eesnäärme puhul välja kommenteeritud
  geom_point(shape=16, size=1.5)+ylim(0, max(andmedAasta$Risk))+
  geom_line(aes(x= Aasta, y = predicted), color = "#D52B1E", linetype = "solid",
size=1)+theme_light()+
  scale_x_continuous(breaks = seq(1970, 2010, by=10),
                    labels = as.character(seq(1970, 2010, by=10)))+
  theme(axis.text.x=element_text(size=12),
        axis.text.y=element_text(size=12))+
  theme(text=element_text(family="serif", size=16))+
  theme(strip.text.x = element_text(size = 20),
        strip.text.y = element_text(size = 20),
        panel.grid.minor = element_blank())

#### JOONIS 2
ggplot(data=andmedAasta, aes(x=Aasta, y=Risk)) +
  facet_grid(Sugu ~ ., scales="free")+ # emakakaela ja eesnäärme puhul välja kommenteeritud
  geom_errorbar(aes(ymin = andmedAasta$alumine,ymax = andmedAasta$ulemine),
               width = 0.5,size = 0.5, color = "grey") +
  geom_point(shape = 16, size = 2)+
  geom_line(aes(X=Aasta, y = predicted), color = "#D52B1E", linetype = "solid", size=1)+
  geom_hline(yintercept=0, color="grey60", linetype = "dashed")+
  scale_x_continuous(breaks = seq(1970, 2010, by=10),
                    labels = as.character(seq(1970, 2010, by=10)))+
  theme(panel.grid.minor = element_blank(),
        axis.ticks = element_blank()+theme_light()+
  theme(axis.text.x=element_text(size=12),
        axis.text.y=element_text(size=12))+

```

```

theme(text=element_text(family="serif", size=16))+
theme(strip.text.x = element_text(size = 20),
      strip.text.y = element_text(size = 20))

#### JOONIS 3
ggplot(visualiseerimiseks, aes(x=Aasta, y=Risk)) +
  facet_grid(rows = vars(Sugu))+ # emakakaela ja eesnäärme puhul välja kommenteeritud
  geom_line() + ylim(0, max(andmedAasta$Risk))+
  geom_point(shape=21, fill=visualiseerimiseks$varvid, size=3)+
  geom_hline(yintercept=0, color="grey60", linetype = "dashed")+
  scale_x_continuous(breaks = seq(1970, 2010, by=10),
                    labels = as.character(seq(1970, 2010, by=10)))+
  theme(axis.text.x=element_text(size=12),
        axis.text.y=element_text(size=12))+
  theme(axis.ticks = element_blank()+theme_light()+
        theme(text=element_text(family="serif", size=16))+
        theme(strip.text.x = element_text(size = 20),
              strip.text.y = element_text(size = 20),
              panel.grid.minor = element_blank()))

```

## Lisa 6. Haigestumise riski standardiseerimine rahvas- tiku soo- ja vanusjaotuse järgi (R)

```

## Andmete ettevalmistamine

### Uurime rahva osakaalusid 2016. aastal vanuse järgi
andmed2016 <- data[data$Aasta == 2016, ]
andmed2016 <- data.frame(Vanus=andmed2016$vanus, Sugu=andmed2016$Sugu,
                        juhtumitearv=andmed2016$juhtumitearv, Rahvaarv=andmed2016$Rahvaarv,
                        Proportsioon=andmed2016$Rahvaarv/sum(andmed2016$Rahvaarv))
andmed2016 <- andmed2016[order(andmed2016$vanus, andmed2016$Sugu),]

### Uurime rahva osakaalusid 1968. aastal vanuse järgi
andmed1968 <- data[data$Aasta == 1968, ]
andmed1968 <- data.frame(Vanus=andmed1968$vanus, Sugu=andmed1968$Sugu,
                        Proportsioon=andmed1968$Rahvaarv/sum(andmed1968$Rahvaarv),
                        juhtumitearv=andmed1968$juhtumitearv, Rahvaarv=andmed1968$Rahvaarv)
andmed1968 <- andmed1968[order(andmed1968$vanus, andmed1968$Sugu),]

### Võrdleme vanusejaotusi
andmed <- rbind(andmed1968, andmed2016)
andmed$Aasta <- "2016"
andmed$Aasta[1:nrow(andmed1968)]<- "1968"
ggplot(andmed, aes(x = Vanus, y = Proportsioon*100, fill = Aasta)) +
  facet_grid(Sugu ~ ., scales="free")+theme_light()+
  geom_bar(data = subset(andmed, Aasta == "1968"), stat = "identity", width = 1, alpha=0.5) +
  geom_bar(data = subset(andmed, Aasta == "2016"), stat = "identity", width = 1, alpha=0.5) +
  scale_x_continuous(breaks = c(seq(0, 85, by=5)),
                    labels = c(as.character(seq(0, 80, by=5)), "85+"))+
  theme(axis.ticks = element_blank()+
        ylab("Osakaal kogu rahvastikust (%)") + xlab("Vanus") +
        theme(axis.text.y = element_text(size = 10))+
        scale_fill_manual("Aasta", values = c("1968"="#7AB800", "2016"="#65CFE9"))+
        theme(panel.grid.minor = element_blank(),
              axis.text.x=element_text(size=12),
              axis.text.y=element_text(size=12))+
        theme(text=element_text(family="serif", size=16),
              legend.text=element_text(size=14))+
        theme(strip.text.x = element_text(size = 20),
              strip.text.y = element_text(size = 20))

## Riskide võrdlemine

### Püüame leida mudeli põhjal prognoosid igale vanusele 1968. aastal
andmestik <- andmed1968
andmestik$Rahvaarv <- ProovRahvaarv
andmestik$Aasta <- 1968
ennustus <- predict(mudel, se.fit=TRUE, newdata=andmestik, type="response")
andmed1968$prognoosid <- ennustus$fit
andmed1968$risk <- ennustus$fit/ProovRahvaarv
andmed1968$standardhalve <- ennustus$se.fit

### Uurime palju keskmiselt oleks olnud haigestunuid mudeli põhjal
haigestunuid1968 <-0
for(i in 1:nrow(andmed1968)){
  haigestunuid1968 <- haigestunuid1968+andmed1968$prognoosid[i]*andmed1968$Proportsioon[i]
}
as.numeric(haigestunuid1968)
sum(andmed1968$juhtumitearv)/sum(andmed1968$Rahvaarv)*ProovRahvaarv

```

```

### Uurime palju keskmiselt oleks olnud haigestunuid mudeli põhjal,
### kui rahva vanuse ja soo osakaalud oleks samad, mis aastal 2016
haigestunuid19682016 <-0
for(i in 1:nrow(andmed1968)){
  haigestunuid19682016 <-
haigestunuid19682016+andmed1968$prognoosid[i]*andmed2016$Proportsioon[i]
}
as.numeric(haigestunuid19682016)

# USALDUSINTERVALL

## 1968. aasta
disp <- 0
for(i in 1:nrow(andmed1968)){
  disp <- disp+(andmed1968$standardhalve[i])**2*(andmed2016$Proportsioon[i])**2
}
(std <- sqrt(disp))
(intervall_ulemine <- haigestunuid19682016+1.96*std)
(intervall_alumine <- haigestunuid19682016-1.96*std)

joonisele <- data.frame(Aastad=1968:2016)
joonisele$SD<-0
joonisele$Tegelik <- 0
joonisele$Standardiseeritud <- 0
for (i in 1968:2016){
  andmed <- data[data$Aasta == i, ]
  andmed <- andmed[order(andmed$Vanus, andmed$Sugu),]
  andmed <- data.frame(Vanus=andmed$Vanus, Sugu=andmed$Sugu,
    Proportsioon=andmed$Rahvaarv/sum(andmed$Rahvaarv),
    juhtumitearv=andmed$juhtumitearv, Rahvaarv=andmed$Rahvaarv)

  andmestik <- andmed
  andmestik$Rahvaarv <- ProovRahvaarv
  andmestik$Aasta <- i
  ennustus <- predict(mudel, se.fit=TRUE, newdata=andmestik, type="response")
  andmed$prognoosid <- ennustus$fit
  andmed$risk <- ennustus$fit/ProovRahvaarv
  andmed$standardhalve <- ennustus$se.fit
  dispersioon <- 0
  haigestunuid <-0
  tegelikult <- 0
  for(j in 1:nrow(andmed)){
    dispersioon <- dispersioon+(andmed$standardhalve[j])**2*(andmed2016$Proportsioon[j])**2
    haigestunuid <- haigestunuid+andmed$prognoosid[j]*(andmed2016$Proportsioon[j])
    tegelikult <- tegelikult+andmed$prognoosid[j]*andmed$Proportsioon[j]
  }
  standardhalve <- sqrt(dispersioon)
  joonisele$Tegelik[i-1967] <- tegelikult
  joonisele$Standardiseeritud[i-1967] <- haigestunuid
  joonisele$SD[i-1967] <- standardhalve
}
joonisele$upr <- joonisele$Standardiseeritud+1.96*joonisele$SD
joonisele$lwr <- joonisele$Standardiseeritud-1.96*joonisele$SD
joonisele$upr; joonisele$lwr

ggplot(data=joonisele, aes(x=Aastad, y=Standardiseeritud)) +
  geom_errorbar(aes(ymin = lwr,ymax = upr),width = 1,size = 0.5, color = "grey") +
  geom_point(shape = 16, size = 1.5, color="black")+
  geom_line(aes(x=Aastad, y = Tegelik), color = "#D52B1E", linetype = "solid", size=1) +
  xlab("Aasta")+
  theme_light() +ylab("keskmise haigestunute arv \n10000 inimese kohta")+
  ylim(0, max(c(joonisele$upr, joonisele$Tegelik)))+
  theme(text=element_text(family="serif", size=16),
    axis.text.x=element_text(size=12),
    axis.text.y=element_text(size=12))

```

## Lisa 7. Prognoosi prognoosivõime uurimine (R)

```

library(stats)
library(gmodels)
library(ggplot2)
options(scipen = 999)
knitr::opts_chunk$set(fig.width=8, fig.height=8)

# Andmete ettevalmistamine
data2017<-aggregate(cbind(juhtumitearv, Rahvaarv)~Vanus + Sugu + Aasta,
  data=koosandmed[(koosandmed$DgnGRnimi == diagnoos & koosandmed$Aasta ==
2017), ], FUN=sum)
if (diagnoos == "Emakakael") {data2017 <- data2017[data2017$Sugu=="Naine", ]}
if(diagnoos == "Eesnääre") {data2017 <- data2017[data2017$Sugu=="Mees", ]}

# Prognoos

```

```

andmestik1 <- data2017[data2017$Sugu=="Mees", c(1,2,3,5)]
andmestik2 <- data2017[data2017$Sugu=="Naine", c(1,2,3,5)]
if (diagnoos == "Emakakael") {andmestik <- andmestik2
} else if(diagnoos == "Eesnääre") {andmestik <- andmestik1
} else {andmestik <- rbind(andmestik1, andmestik2)}
data2017$predicted <- exp(predict(mudel, andmestik, type="link"))
data2017$ulemine <- 0; data2017$alumine <- 0
for (i in 1:nrow(data2017)){
  test<-binom.test(data2017$juhtumitearv[i], data2017$Rahvaarv[i])
  data2017$alumine[i] <- test$conf.int[1]*data2017$Rahvaarv[i]
  data2017$ulemine[i] <- test$conf.int[2]*data2017$Rahvaarv[i]
}
ggplot(data2017, aes(x=vanus, y=juhtumitearv))+
  facet_grid(Sugu ~ ., scales="free")+ # emakakaela ja eesnäärme puhul välja kommenteeritud
  geom_errorbar(aes(ymin = alumine,ymax = ulemine),width = 0.5,size = 0.5, color = "grey") +
  geom_point(shape=21, fill="black", color="black", size=2)+theme_light()+
  geom_line(aes(X=vanus, y=predicted), color="red", size=1) + ylab("Vähijuhtude arv")+
  geom_hline(yintercept=0, color="grey70", linetype = "dashed")+
  scale_x_continuous(breaks = seq(0, 85, by=5),
                    labels = c(as.character(seq(0, 80, by=5)), "85+"))+
  theme(panel.grid.minor = element_blank(),
        axis.text.x=element_text(size=12),
        axis.text.y=element_text(size=12))+
  theme(text=element_text(family="serif", size=16))+
  theme(strip.text.x = element_text(size = 20),
        strip.text.y = element_text(size = 20))

```

## Lisa 8. Mudeli $s(x,y,z)$ liikme kuju kõikide paikmete korral

$$\begin{aligned}
s(x,y,z) = & 39240315.67047 - 31649642.29783 \cdot I(z = \text{Naine}) \\
& + 0.15168 \cdot x - 0.07014x^2 + 0.00493 \cdot x^3 \\
& - 0.00677 \cdot (x - 7)^3 \cdot I(x - 7 \geq 0) + 0.00252 \cdot (x - 15)^3 \cdot I(x - 15 \geq 0) \\
& - 0.00092 \cdot (x - 23)^3 \cdot I(x - 23 \geq 0) + 0.00031 \cdot (x - 31)^3 \cdot I(x - 31 \geq 0) \\
& - 0.00025 \cdot (x - 39)^3 \cdot I(x - 39 \geq 0) + 0.00011 \cdot (x - 46)^3 \cdot I(x - 46 \geq 0) \\
& + 0.00007 \cdot (x - 54)^3 \cdot I(x - 54 \geq 0) - 0.00005 \cdot (x - 62)^3 \cdot I(x - 62 \geq 0) \\
& + 0.00005 \cdot (x - 70)^3 \cdot I(x - 70 \geq 0) - 0.00007 \cdot (x - 78)^3 \cdot I(x - 78 \geq 0) \\
& - 59713.77059 \cdot y + 30.28970 \cdot y^2 - 0.00512 \cdot y^3 \\
& + 0.00591 \cdot (y - 1972)^3 \cdot I(y - 1972 \geq 0) - 0.00094 \cdot (y - 1977)^3 \cdot I(y - 1977 \geq 0) \\
& + 0.00004 \cdot (y - 1982)^3 \cdot I(y - 1982 \geq 0) + 0.00026 \cdot (y - 1987)^3 \cdot I(y - 1987 \geq 0) \\
& - 0.00054 \cdot (y - 1992)^3 \cdot I(y - 1992 \geq 0) + 0.00108 \cdot (y - 1997)^3 \cdot I(y - 1997 \geq 0) \\
& - 0.00126 \cdot (y - 2002)^3 \cdot I(y - 2002 \geq 0) + 0.00057 \cdot (y - 2007)^3 \cdot I(y - 2007 \geq 0) \\
& + 0.00048 \cdot (y - 2012)^3 \cdot I(y - 2012 \geq 0) \\
& - 0.00370 \cdot I(z = \text{Naine}) \cdot x + 0.00461 \cdot I(z = \text{Naine}) \cdot x^2 - 0.00038 \cdot I(z = \text{Naine}) \cdot x^3 \\
& + 0.00070 \cdot I(z = \text{Naine}) \cdot (x - 7)^3 \cdot I(x - 7 \geq 0) \\
& - 0.00046 \cdot I(z = \text{Naine}) \cdot (x - 15)^3 \cdot I(x - 15 \geq 0) \\
& - 0.00010 \cdot I(z = \text{Naine}) \cdot (x - 23)^3 \cdot I(x - 23 \geq 0) \\
& + 0.00037 \cdot I(z = \text{Naine}) \cdot (x - 31)^3 \cdot I(x - 31 \geq 0) \\
& - 0.00025 \cdot I(z = \text{Naine}) \cdot (x - 39)^3 \cdot I(x - 39 \geq 0) \\
& + 0.00036 \cdot I(z = \text{Naine}) \cdot (x - 46)^3 \cdot I(x - 46 \geq 0) \\
& - 0.00030 \cdot I(z = \text{Naine}) \cdot (x - 54)^3 \cdot I(x - 54 \geq 0) \\
& + 0.00008 \cdot I(z = \text{Naine}) \cdot (x - 62)^3 \cdot I(x - 62 \geq 0) \\
& - 0.00012 \cdot I(z = \text{Naine}) \cdot (x - 70)^3 \cdot I(x - 70 \geq 0) \\
& + 0.00014 \cdot I(z = \text{Naine}) \cdot (x - 78)^3 \cdot I(x - 78 \geq 0) \\
& + 48163.12142 \cdot I(z = \text{Naine}) \cdot y - 24.43087 \cdot I(z = \text{Naine}) \cdot y^2 \\
& + 0.00413 \cdot I(z = \text{Naine}) \cdot y^3 \\
& - 0.00485 \cdot I(z = \text{Naine}) \cdot (y - 1972)^3 \cdot I(y - 1972 \geq 0) \\
& + 0.00113 \cdot I(z = \text{Naine}) \cdot (y - 1972)^3 \cdot I(y - 1972 \geq 0)
\end{aligned}$$

$$\begin{aligned}
& -0.00077 \cdot I(z = \text{Naine}) \cdot (y - 1977)^3 \cdot I(y - 1977 \geq 0) \\
& +0.00087 \cdot I(z = \text{Naine}) \cdot (y - 1982)^3 \cdot I(y - 1982 \geq 0) \\
& -0.00102 \cdot I(z = \text{Naine}) \cdot (y - 1987)^3 \cdot I(y - 1987 \geq 0) \\
& +0.00055 \cdot I(z = \text{Naine}) \cdot (y - 1992)^3 \cdot I(y - 1992 \geq 0) \\
& +0.00021 \cdot I(z = \text{Naine}) \cdot (y - 1997)^3 \cdot I(y - 1997 \geq 0) \\
& +0.00007 \cdot I(z = \text{Naine}) \cdot (y - 2002)^3 \cdot I(y - 2002 \geq 0) \\
& -0.00285 \cdot I(z = \text{Naine}) \cdot (y - 2007)^3 \cdot I(y - 2007 \geq 0)
\end{aligned}$$

## Lisa 9. Mudeli $s(x, y, z)$ liikme kuju emakakaelavähi korral

$$\begin{aligned}
s(x, y, z) = & -1150481.77576 + 3.94943 \cdot x - 0.12872x^2 + 0.00146 \cdot x^3 \\
& -0.00140 \cdot (x - 28)^3 \cdot I(x - 28 \geq 0) - 0.00009 \cdot (x - 57)^3 \cdot I(x - 57 \geq 0) \\
& +1748.22966 \cdot y - 0.88554 \cdot y^2 + 0.00015 \cdot y^3 \\
& -0.00027 \cdot (y - 1980)^3 \cdot I(y - 1980 \geq 0) + 0.00026 \cdot (y - 1992)^3 \cdot I(y - 1992 \geq 0) \\
& -0.00066 \cdot (y - 2004)^3 \cdot I(y - 2004 \geq 0)
\end{aligned}$$

## Lisa 10. Mudeli $s(x, y, z)$ liikme kuju eesnäärmevähi korral

$$\begin{aligned}
s(x, y, z) = & 864949.79085 - 0.80013 \cdot x + 0.03405 \cdot x^2 - 0.00032 \cdot x^3 \\
& +0.00035 \cdot (x - 42.5)^3 \cdot I(x - 42.5 \geq 0) \\
& -1308.79555 \cdot y + 0.66010 \cdot y^2 - 0.00011 \cdot y^3 \\
& +0.00018 \cdot (y - 1984)^3 \cdot I(y - 1984 \geq 0) - 0.00041 \cdot (y - 2000)^3 \cdot I(y - 2000 \geq 0)
\end{aligned}$$



## Tervisestatistilised andmed ja informatsioon:

- **Tervisestatistika ja terviseuuringute andmebaas**  
<http://www.tai.ee/tstua>
- **Tervise Arengu Instituudi tervisestatistika osakonna veebileht**  
<http://www.tai.ee/tegevused/tervisestatistika>
- **Andmepäring Tervise Arengu Instituudile**  
[tai@tai.ee](mailto:tai@tai.ee)
- **Statistikaameti andmebaas**  
<http://www.stat.ee/>
- **Euroopa Liidu statistika**  
<http://ec.europa.eu/eurostat>
- **Maailma Terviseorganisatsiooni (WHO) Euroopa esinduse tervise andmebaas**  
<http://data.euro.who.int/hfad/>
- **Majandusliku Koostöö ja Arengu Organisatsiooni (OECD) andmebaas**  
[http://stats.oecd.org/index.aspx?DataSetCode=HEALTH\\_STAT](http://stats.oecd.org/index.aspx?DataSetCode=HEALTH_STAT)

