

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Karmel Teder

**Absoluutriski prognoosimine konkureerivate
riskide korral: meetodite võrdlus teise tüüpi
diabeedi näitel**

Matemaatika ja statistika õppekava

Matemaatilise statistika eriala

Magistritöö (30 EAP)

Juhendaja: prof. Krista Fischer, PhD

TARTU 2025

**ABSOLUUTRISKI PROGNOOSIMINE KONKUREERIVATE
RISKIDE KORRAL: MEETODITE VÕRDLUS TEISE TÜÜBI
DIABEEDI NÄITEL**

Magistritöö

Karmel Teder

Lühikokkuvõte

Käesoleva magistritöö eesmärk on anda ülevaade konkureerivaid riske arvesse võtvatest haigusrisiki hindamise meetoditest ning võrrelda neist peamiseid Tartu Ülikooli Eesti geenivaramu andmetel teise tüüpi diabeedi riski hindamisel. Peamiselt soovitakse näha, millise mudeliga saaks kõige paremini hinnata vanemaealiste inimeste teist tüüpi diabeeti haigestumise riski. Töös võrreldakse omavahel Coxi võrdeliste riskide mudelit, Fine-Gray võrdeliste alamriskide mudelit ning pseudovaatluste abil hinnatud lineaarset riskimudelit. Mudelite abil hinnatakse teise tüüpi diabeeti haigestumise 10 aasta absoluutriski, võttes seejuures arvesse võimalust enne haigestumist surra. Töö tulemusena selgub, et kõige täpsemalt võimaldab riski hinnata konkureerivaid riske arvesse võttev Coxi võrdeliste riskide mudel.

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

Märksõnad: elukestusanalüüs, geenivaramu, polügeenne riskiskoor, teist tüüpi diabeet, konkureerivad riskid.

**ABSOLUTE RISK PREDICTION IN THE PRESENCE OF
COMPETING RISKS: A COMPARISON OF METHODS USING
THE EXAMPLE OF TYPE 2 DIABETES**

Master's thesis

Karmel Teder

Abstract

The aim of this master's thesis is to provide an overview of disease risk prediction methods that account for competing risks and to compare the main approaches using data from the Estonian Biobank at the University of Tartu for assessing the risk of type 2 diabetes. The goal is to determine which model estimates the risk of developing type 2 diabetes among older individuals the best. The thesis compares Cox proportional hazards model, Fine-Gray subdistribution hazards model, and a linear risk model estimated using pseudovalues. The models are used to estimate the risk of developing type 2 diabetes while accounting for the competing risk of death prior to disease onset. As a result it is seen that the most accurate risk estimates are provided by Cox proportional hazards model accounting for competing risks.

CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematics.

Key Words: survival analysis, biobank, polygenic risk score, type two diabetes, competing risks.

Sisukord

| | |
|---|-----------|
| Sissejuhatus | 4 |
| 1 Statistilised meetodid | 6 |
| 1.1 Haigestumise absoluutrisi defineerimine | 6 |
| 1.2 Riski hindamine konkureerivate riskide korral | 9 |
| 1.2.1 Coxi võrdeliste riskide mudel | 10 |
| 1.2.2 Fine ja Gray mudel | 13 |
| 1.3 Coxi ja Fine-Gray mudeli võrdlus simulatsiooni abil | 16 |
| 1.4 Tõkestatud keskmise eluea hinnang | 18 |
| 2 60—90-aastaste geenidoonorite teise tüüpi diabeedi riski hinnangud | 22 |
| 2.1 Ülevaade andmetest | 22 |
| 2.2 Analüüsi käik | 24 |
| 2.3 Tulemused | 26 |
| Kokkuvõte | 36 |
| Kasutatud allikad | 38 |

Sissejuhatus

Käesoleva magistritöö eesmärk on anda ülevaade konkureerivaid riske arvesse võtvatest haigusrisiki hindamise meetoditest ning võrrelda neist peamiseid Tartu Ülikooli Eesti geenivaramu andmetel teise tüüpi diabeedi (T2D) riski hindamisel. Eelkõige on soov välja selgitada, millise mudeliga saaks kõige paremini hinnata vanemaealiste inimeste teist tüüpi diabeeti haigestumise riski. Teise tüüpi diabeet on üks olulisemaid probleeme rahvatervishoius. Maailma Terviseorganisatsiooni andmetel oli aastal 2022 ligi 14% maailma täiskasvanud elanikkonnast mõni diabeetivorm ning diabeedi levimus on maailmas kasvutrendis. Seejuures enam kui 95% diabeetikutest on teise tüüpi diabeet. (World Health Organization, 2024)

Teise tüüpi diabeet on aga haigus, mida saab tervislike eluviisidega ennetada või selle avaldumist edasi lükata. Näiteks vähendavad haigusrisiki tasakaalustatud toitumine, rohke liikumine, normaalkaalulisus ning tubaka kasutamisest loobumine. Lisaks haiguse ennetamisele, aitavad need elustiilitegurid kaasa ka haiguse kontrolli all hoidmisele. See on oluline, kuna diabeet võib aja jooksul mõjutada närve ja veresooni ning põhjustada näiteks pimedaks jäämist, neerupuudulikkust ja südamehaigusi. (World Health Organization, 2024) Siiski mõjutab teise tüüpi diabeedi avaldumist ka pärilikkus (Prasad ja Groop, 2015). Seetõttu on oluline, et teise tüüpi diabeedi ennetava sekkumise planeerimisel ning inimeste nõustamisel nende haigusrisiki osas oleks võimalik haigusrisiki võimalikult hästi hinnata, võttes seejuures arvesse nii geneetilisi kui elustiiliga seotud riskifaktoreid.

Haigestumise riski hindamisel on oluline arvesse võtta, et mitte kõik inimesed ei haigestu uuritavasse haigusesse ning võib juhtuda, et toimub sündmus, mis välistab sellesse haigusesse haigestumise. Selliseid sündmusi, mis uuritava sündmuse toimumise välistavad, nimetatakse konkureerivateks riskideks (Collett, 2015). Teise tüüpi diabeedi puhul on konkureerivad riskid enim probleemiks vanemaealiste inimeste haigetumisrisiki hindamisel, kui arvesse tuleks võtta võimalust, et inimene sureb

mõnel muul põhjusel kui diabeet.

Haigestumise riski hindamiseks viisil, mis võtab arvesse ka konkureerivate riskide olemasolu, on olemas mitmeid võimalusi. Siiski pole tihti selge, millist neist tuleks eelistada. (Conner *et al.*, 2022; Allison, 2018) Seetõttu soovitakse käesolevas magistritöös võrrelda, kuidas erinevad erinevate meetoditega hinnatud Tartu Ülikooli Eesti geenivaramu 60—90-aastaste geenidoonorite teise tüübi diabeeti haigestumise riski prognoosid.

Töö koosneb kahest peatükist. Neist esimeses on kirjeldatud kasutatavaid statistilisi meetodeid ning see jaguneb neljaks alapeatükiks, milles kirjeldatakse haigusrisiki ning selle hindamise meetodeid. Ühes esimese peatüki alapeatükkidest kirjeldatakse ka simulatsiooni kahe konkureerivaid riske arvesse võtva meetodi võrdlemiseks. Magistritöö teine peatükk koosneb kolmest alapeatükist, milles antakse ülevaade töös kasutatud andmetest ning nendega tehtud analüüsist ja võetakse kokku töö käigus saadud tulemused.

1 Statistilised meetodid

Elukestusanalüüsis on konkureerivate riskide arvesse võtmiseks olemas erinevaid meetodeid. Käesolevas peatükis on elukestusanalüüsis konkureerivaid riske arvesse võtvate meetoditena kirjeldatud Coxi võrdeliste riskide meetodit, Fine ja Gray võrdeliste alamiriskide meetodit ning tõkestatud keskmise eluea meetodit. Peatükk on jagatud neljaks alapeatükiks. Neist esimeses on kirjeldatud, kuidas defineeritakse haigusrisi ilma konkureerivate riskideta ja ka seda, kuidas on seda võimalik teha konkureerivate riskide korral. Teises alapeatükis on kirjeldatud Coxi ning Fine-Gray meetodeid riski hindamiseks konkureerivate riskide korral. Kolmandas alapeatükis on kirjeldatud simulatsiooni nende meetodite võrdlemiseks. Viimases alapeatükis on kirjeldatud tõkestatud keskmise eluea hinnangut alternatiivina haigusrisi hindamiseks.

Käesoleva peatüki kaks esimest alapeatükki on kirjutatud Collett (2015) põhjal, kui ei ole märgitud teisiti.

1.1 Haigestumise absoluutriski defineerimine

Haigestumise absoluutriskina vaadeldakse selle haiguse kumulatiivset esinemisfunktsiooni kindlal ajahetkel t^* (Conner *et al.*, 2022). Tähistagu T juhuslikku suurust, mis vastab ajale jälgimisaja algusest haigestumiseni ning olgu selle juhusliku suuruse tihedusfunktsioon f . Üldjuhul on kumulatiivne esinemisfunktsioon juhusliku suuruse T jaotusfunktsioon ehk

$$F(t^*) = P(T < t^*) = \int_0^{t^*} f(t)dt.$$

On selge, et nii on risk üldse kunagi haigestuda 1. Reaalses elus võib aga enne haigestumist saabuda surm või mõni muu sündmus, mis välistab haigestumise uuritavasse haigusesse. Seetõttu on elulemusanalüüsis vajalik võtta arvesse konkureerivaid riske.

Konkureerivate riskide puudumisel vaadeldakse elulemusanalüüsis elulemust T kui aega katseisiku vaatlusperioodi algusest tema haigestumiseni. Võrdelisi riske arvesse võtva analüüsi puhul võetakse riski hindamisel arvesse teisi sündmusi, mis huvialuse haiguse saamise välistavad või edasi lükkavad. Näiteks konkreetse haigusesse suremise riski analüüsid võivad konkureerivad sündmused olla mingisse teise haigusesse suremine ja surm mõne välise põhjuse tõttu. Sellisel juhul saame defineerida m juhuslikku suurust T_1, T_2, \dots, T_m , mis vastavad m võimaliku konkureeriva sündmuse toimumise aegadele. Reaalsuses on vaadeldav vaid esimesena toimuv sündmus ehk $T = \min(T_1, T_2, \dots, T_m)$. Seega on konkreetse indiviidi puhul vaadeldavad juhuslikud suurused (T, C) , kus $C \in \{1, 2, \dots, m\}$ tähistab seda, milline konkureerivatest sündmustest toimus.

Käesoleva magistritöö praktilises osas arvutatakse teise tüüpi diabeeti haigestumise riski vanemaealistele inimestele ning seega saab C saada väärtuse $c = 1$, kui katseisiku vaatlusperiood lõppeb T2D diagnoosiga, ning $c = 2$, kui katseisiku vaatlusperiood lõppeb tema surmaga ilma T2D diagnoosita. Ollakse huvitatud T2D absoluutriskist, võttes surma arvesse konkureeriva riskina.

Olukorras, kus analüüsitakse ühe sündmuse absoluutriski võttes konkureerivate riskidena arvesse ka teisi sündmusi, vaadeldakse selle sündmuse põhjusespetsiifilist esinemusfunktsiooni. Haigusrisiki analüüsimise korral nimetatakse seda funktsiooni ka põhjusjärgseks või põhjusespetsiifiliseks haigestumuseks. Ajahetkel t on põhjusespetsiifiline esinemusfunktsioon defineeritud kui

$$F_j(t) = P(T < t, C = j),$$

kus $j = 1, 2, \dots, m$.

Põhjusjärgse haigestumuse korral on risk tõkestatud enamasti ühest väiksema suu-

rusega. Seda seetõttu, et

$$F_j(\infty) = P(T < \infty, C = j) = P(C = j).$$

Seega on põhjusjärgse haigestumuse analüüsimisel haigestumusfunktsiooni näol tegemist mittetäieliku jaotusfunktsiooniga elulemusele.

Elukestusanalüüsis pakub lisaks huvi ka riskifunktsioon. Juhul, kui konkureerivaid riske ei esine, avaldub see kujul

$$h(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T < t + \delta t \mid T \geq t)}{\delta t},$$

kusjuures saab näidata, et see on võrdne elulemusandmete tihedusfunktsiooni ning elulemusfunktsiooni jagatisega. Konkureerivate riskide puhul on j -nda põhjuse järgne riskifunktsioon defineeritud kui

$$h_j(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T < t + \delta t, C = j \mid T \geq t)}{\delta t}, \quad (1)$$

mille puhul saab samuti näidata, et see on võrdne põhjusespetsiifilise elukestusandmete tihedusfunktsiooni $f_j(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T < t + \delta t, C = j)}{\delta t}$ ning üldmise elulemusfunktsiooni jagatisega ehk

$$h_j(t) = \frac{f_j(t)}{S(t)},$$

kus $S(t) = P(T \geq t)$. Seega saame põhjusjärgse haigestumusfunktsiooni $F_j(t)$ avaldada kui

$$F_j(t) = \int_0^t h_j(u) S(u) du. \quad (2)$$

Põhjusjärgse haigestumusfunktsiooni hindamiseks on erinevaid meetodeid, millest mõnda tutvustatakse järgnevas alapeatükis. Seejuures on üldjuhul haiguse absoluutrisi hindamise meetodite eelduseks elulemusaegade sõltumatus tsenseerimisaegadest.

1.2 Riski hindamine konkureerivate riskide korral

Huvipakkuvate parameetrite hindamine elukestusanalüüsis on keeruline tsenseerimise tõttu. Tsenseerimiseks nimetatakse seda, kui indiviidi jälgimisaeg lõppeb, kuid uuritavat sündmust (haigestumist) pole veel toimunud. Seega, kui i -nda indiviidi vaatlusaeg lõppeb ajahetkel t_i , siis võib selle põhjuseks olla kas üks konkureerivatest sündmustest või tsenseerimine.

Vaatleme esmalt Kaplan-Meieri hinnangut, mis ignoreerib konkureerivaid riske ehk käsitleb konkureerivaid sündmusi tsenseerimisena. Olgu $t_1 < t_2 < \dots < t_n$ ajahetked, kus toimub huvipakkuv sündmus või tsenseerimine, kusjuures eeldame lihtsuse huvides, et ühel ajahetkel saab saab sündmus või tsenseerimine toimuda vaid ühel inimesel. Olgu $Y(t) = \sum_{i=1}^n I(t_i \geq t)$ nende katseisikute arv, kellel enne ajahetke t huvialust sündmust ega tsenseerimist toimunud ei ole ning $\Delta N_j(t_i)$ nende katseisikute arv, kellel toimub ajahetkel t_i huvialune sündmus ehk kui $N_j(t) = \sum_{i=1}^n I(t_i \leq t, C = j)$, siis $\Delta N_j(t_i) = N_j(t_i) - N_j(t_{i-1})$. Kaplan-Meieri hinnang elulemusfunktsioonile $S_j(t)$ avaldub siis kui

$$\hat{S}_j(t) = \prod_{i:t_i \leq t} \left(1 - \frac{\Delta N_j(t_i)}{Y(t_i)} \right).$$

See on aga nihkega hinnang kumulatiivsele haigestumusele, sest $Y(t)$ kahaneb ajas mitte ainult tsenseerimiste, vaid ka konkureerivate sündmuste tõttu.

Vaatleme näiteks olukorda, kus algselt on valimis 10 isikut ning 10 aasta jooksul neist 2 haigestub diabeeti ja 3 sureb mingil muul põhjusel. Seega on 10 aasta diabeedirisk selles valimis $\frac{2}{10} = 0,2$. Oletame aga, et esimene isik haigestub diabeeti ajahetkel t_1 , kui selleks hetkeks on üks isik valimist juba surnud, ning teine isik haigestub ajahetkel t_2 , kui kaks isikut on juba surnud. Seega $Y(t_1) = 9$ ning $Y(t_2) = 7$ ning ülaltoodud valemi põhjal saame $S(10)$ hinnanguks $(1 - \frac{1}{9}) \cdot (1 - \frac{1}{7}) = 0,762$ ning esinemusfunktsiooni hinnanguks $1 - 0,762 = 0,238$, mis on tegelikku riski selgelt üle hinnanud. Kui siin näites hindasime alguses riski klassikalise tõenäosuse

põhimõttel, siis reaalsuses ei ole see aga võimalik tsenseerimise tõttu. On näidatud, et riski hindamiseks konkureerivate riskide korral sobib Aalen-Johanseni hinnang (Hansen *et al.*, 2017).

Kaplan-Meieri hinnanguid saame kasutada Aalen-Johanseni hinnangute avaldamiseks. Aalen-Johanseni mitteparameetrilise hinnangu abil saame põhjusjärgset haigestumusfunktsiooni konkureerivate riskide korral. Hinnangu leidmiseks on esmalt vaja leida Kaplan-Meieri hinnang üleüldisele elulemusfunktsioonile $S(t)$, mis avaldub kui

$$\hat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{\Delta N(t_i)}{Y(t_i)} \right),$$

kus $\Delta N(t_i)$ on nende isikute arv, kellel toimub ükskõik milline m sündmusest ajahetkel t_i . Siis Aalen-Johanseni hinnang põhjusjärgsele haigestumusfunktsioonile põhineb avaldisel (2) ja avaldub kui

$$\hat{F}_j(t) = \sum_{i:t_i \leq t} \hat{S}(t_{i-1}) \frac{\Delta N_j(t_i)}{Y(t_i)}.$$

Seejuures $N_j(t) = \sum_{i=1}^n I(T \leq t, C = j)$. (Conner *et al.*, 2022; Hansen *et al.*, 2017)

Praktikas kasutatakse Aalen-Johanseni hinnanguid ka pseudovaatluste tekitamiseks, et saaks kasutada tavapäraseid statistilisi meetodeid, mida muidu elukestusandmetel nende ebasümmeetrilise jaotuse ning tsenseerimise tõttu kasutada ei saa (Conner *et al.*, 2022; Andersen ja Pohar Perme, 2010). Täpsemalt on seda kirjeldatud käesoleva peatüki kolmandas alapeatükis.

1.2.1 Coxi võrdeliste riskide mudel

Poolparameetriselt saab haigestumusriski hinnata näiteks Coxi mitme seisundiga mudelitega. Juhul, kui konkureerivaid riske ei esine, kasutatakse riski hindamiseks Coxi võrdeliste riskide mudelit. Võrdeliste riskide mudeli puhul eeldatakse, et võrreldes kahte gruppi, on igal ajahetkel riskifunktsiooni väärtus ühes grupis

proportsionaalne teise grupi sarnaste omadustega katsealuse riskiga. Tähistagu \mathbf{x}_i vektorit i -ndat katsealust kirjeldavate tunnustega ning h_0 baasriskifunktsiooni ehk riskifunktsiooni juhul, kui kirjeldavate tunnuste vektor on võrdne nullvektoriga. Võrdeliste riskide mudel riskifunktsioonile i -nda indiviidi jaoks avaldub kui

$$h_i(t^*) = h_0(t^*) \exp(\beta' \mathbf{x}_i),$$

kus β on vektor kirjeldavate tunnuste kordajatega ning $'$ tähistab transponeerimist. Selle abil avaldub kumulatiivne esinemusfunktsioon kui

$$\hat{F}(t^*) = 1 - \exp\left(-H_0(t^*)e^{\beta' \mathbf{x}_i}\right),$$

kus $H_0(t^*) = \int_0^{t^*} h_0(t)dt$ on kumulatiivne baasriskifunktsioon. (Therneau, Crowson ja Atkinson, 2024; Collett, 2015)

Coxi võrdeliste riskide mudeli parameetrid β hinnatakse osalise tõepära meetodil nii, et baasriskifunktsiooni $h_0(t)$ kohe ei hinnata. Olgu t_i ($i = 1, 2, \dots, n$) n indiviidi elukestused ning δ_i indikaatortunnus, mis võtab väärtuse 1, kui ajahetkel t_i toimus i -ndal katsealusel huvipakkuv sündmus ning väärtuse 0 muudel juhtudel. Tähistagu $R(t_i)$ riskihulka ehk nende katsealuste indeksite hulka, kellel pole enne ajahetke t_i huvialust sündmust ega tsenseerimist toimunud. Siis osalise tõepära funktsioon Coxi võrdeliste riskide mudelile avaldub kui

$$L = \prod_{i=1}^n \left(\frac{h_0(t_i) \exp(\beta' \mathbf{x}_i)}{\sum_{l \in R(t_i)} h_0(t_i) \exp(\beta' \mathbf{x}_l)} \right)^{\delta_i} = \prod_{i=1}^n \left(\frac{\exp(\beta' \mathbf{x}_i)}{\sum_{l \in R(t_i)} \exp(\beta' \mathbf{x}_l)} \right)^{\delta_i}.$$

Parameetrite hinnangud saadakse selle logaritmi maksimeerimisel. (Cox ja Oakes, 2018; Collett, 2015)

Pärast β hindamist on võimalik hinnata kumulatiivset baasriskifunktsiooni. Selleks saab kasutada Breslow meetodit, mis on tuntud ka Nelson-Aaleni hinnanguna ku-

mulatiivsele baasriskifunktsioonile. Breslow hinnang kumulatiivsele baasriskifunktsioonile on

$$\hat{H}_0(t^*) = \sum_{t \leq t^*} \frac{\Delta N(t)}{\sum_{l \in R(t)} \exp(\hat{\beta}' \mathbf{x}_l)},$$

kus $\Delta N(t)$ tähistab nende katseisikute arvu, kellel toimub ajahetkel t huvialune sündmus.

Konkureerivate riskide korral kasutatakse põhjusjärgsete absoluutriskide hindamiseks põhjusespetsiifilisi riskifunktsioone. Eeldades võrdeliste riskide mudelit, avaldub j -nda põhjuse järgne riskifunktsioon i -ndale katsealusele kui

$$h_{ij}(t^*) = h_{0j}(t^*) \exp(\beta_j' \mathbf{x}_i),$$

kus h_{0j} tähistab põhjuse j spetsiifilist baasriskifunktsiooni ning β_j vektorile \mathbf{x} vastavate tunnuste kordajate vektorit. Põhjusjärgne haigestumus avaldub põhjusjärgsete riskifunktsioonide kaudu kui

$$F_{ij}(t^*) = \int_0^{t^*} h_{ij}(t) S_i(t) dt,$$

kus

$$S_i(t^*) = \exp\left(\sum_{j=1}^m H_{ij}(t^*)\right).$$

Seejuures $H_{ij}(t^*) = \int_0^{t^*} h_{ij}(t) dt$. (Collett, 2015; Benichou ja Gail, 1990)

Olgu δ_{ij} indikaator-tunnus väärtusega 1, kui i -ndal katsealusel toimub sündmus, mille korral $C = j$ ning väärtusega 0 muudel juhtudel. Põhjusjärgse Coxi mudeli tõepärafunktsioon avaldub kui

$$L = \prod_{i=1}^n \prod_{j=1}^m \left(\frac{\exp(\beta_j' \mathbf{x}_i)}{\sum_{l \in R(t_i)} \exp(\beta_l' \mathbf{x}_i)} \right)^{\delta_{ij}}.$$

Siit on näha, et j -nda põhjuse järgne osaline tõepärafunktsioon on

$$L_j = \prod_{i=1}^n \left(\frac{\exp(\beta_j' \mathbf{x}_i)}{\sum_{l \in R(t_i)} \exp(\beta_l' \mathbf{x}_i)} \right)^{\delta_{ij}},$$

mille maksimeerimisel on võimalik hinnata parameetrite β_j väärtuseid ehk seda, kuidas mõjutavad kirjeldavad tunnused j -nda sündmuse toimumise riski. Seega hinnatakse konkureerivate riskide korral m tavalist Coxi mudelit, kus konkureerivaid sündmuseid käsitletakse tsenseerimisena.

Juhul, kui konkureerivaid riske ei esine, on riski võimalik hinnata ka näiteks Weibulli parameetrilise mudeliga. Seda mudelit kasutas autor T2D 10 aasta haigestumise riski hindamiseks oma bakalaureusetöös. (Teder, 2023) Teoreetiliselt saab haigestumise absoluutriski Weibulli mudelite abil hinnata ka konkureerivate riskide esinedes. Praktikas on see aga keeruline, mistõttu Weibulli mudelite kasutamist absoluutriski hindamiseks konkureerivate riskide korral käesolevas magistris ei vaadelda.

1.2.2 Fine ja Gray mudel

Põhjusjärgset kumulatiivset haigestumust saab hinnata ka Fine-Gray mudeliga. See on mudel, mille abil hinnatakse riskitegurite mõju alamjaotuse riskifunktsioonile (inglise keeles *subdistribution hazard*) ehk alamriskile, mis on defineeritud kui

$$\lambda_j(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T < t + \delta t, C = j \mid T \geq t \cup (T \leq t \cap C \neq j))}{\delta t}. \quad (3)$$

Paneme tähele, et põhjusjärgsetest riskidest erinevad alamriskid selle poolest, et põhjusjärgse riski avaldise (1) lugejas oleva tõenäosuse tingimuseks on, et toimunud ei ole ühtegi sündmust, kuid alamriski (3) puhul on vastavaks tingimuseks, et toimunud ei ole vaadeldavat sündmust. (Collett, 2015; Conner *et al.*, 2022)

Fine-Gray riskimudel alamriski hindamiseks põhjuse, mille korral $C = j$, korral katsealusele tunnustevektoriga \mathbf{x}_i on

$$\lambda_{ij}(t) = \lambda_{0j}(t^*) \exp(\beta_j' \mathbf{x}_i),$$

kus $\lambda_{0j}(t)$ tähistab j -nda põhjuse järgset baasalamriski ajahetkel t ning β_j mudelikordajaid. Selle mudeli abil hinnatakse põhjusjärgne kumulatiivne haigestumus kui

$$\hat{F}_{ij}(t^*) = 1 - \exp\left(-\Lambda_{0j}(t^*) e^{\beta_j' \mathbf{x}_i}\right).$$

Seejuures $\Lambda_{0j}(t)$ tähistab kumulatiivset baasalamriski, mis on saadud riskifunktsiooni integreerides. (Conner *et al.*, 2022; Fine ja Gray, 1999; Collett, 2015)

Sarnaselt Coxi võrdeliste riskide mudelile, eeldab ka Fine-Gray mudel riskide võrdelisust, kuid Fine-Gray mudel eeldab alamriskide võrdelisust (Collett, 2015; Conner *et al.*, 2022). Mudelikordajad hinnatakse ka Fine-Gray mudelis osalise tõepära meetodil, kuid riskihulga väärtused kaalutakse kordajaga w_{il} . Tähitagu δ_{ij} jätkuvalt indikaatorit, kas i -ndal katseisikul toimub sündmus, mille korral $C = j$. Siis j -nda põhjuse järgne osaline tõepärafunktsioon on Fine ja Gray mudeli puhul

$$L_j = \prod_{i=1}^n \left(\frac{\exp(\beta_j' \mathbf{x}_i)}{\sum_{l \in R(t_i)} w_{il} \exp(\beta_l' \mathbf{x}_i)} \right)^{\delta_{ij}},$$

kus n on katseisikute arv ja $R(t)$ riskihulk ajahetkel t .

Tähistagu $\hat{S}_c(t_i)$ Kaplan-Meieri hinnangut tsenseerimise elulemusfunktsioonile ehk funktsioonile $P(Z > t_i)$, kus Z on indiviidi tsenseerimise aeg, ajahetkel t_i . Siis kaalud w_{il} , mida hinnangus kasutatakse, arvutatakse kui

$$w_{il} = \frac{\hat{S}_c(t_i)}{\hat{S}_c(\min\{t_i, t_l\})}.$$

Hinnangus tähistab t_l l -nda katsealuse elulemust.

Sarnaselt Coxi mudelile hinnatakse ka Fine-Gray mudelis kumulatiivne baasriskifunktsioon Breslow hinnanguga. Erinevalt Coxi mudelist, kasutatakse Fine ja Gray mudeli puhul kaalutud hinnangut

$$\hat{\Lambda}_{0j}(t^*) = \sum_{t_i \leq t^*} \frac{\Delta N(t_i)}{\sum_{l \in R(t_i)} w_{il} \exp(\beta_j' \mathbf{x}_l)},$$

kus w_{il} tähistavad kaalusid ning $R(t^*)$ on nende inimeste riskihulk, kellel ei ole huvialust sündmust toimunud, kuid võib olla toimunud mõni konkureerivatest sündmustest. $\Delta N(t)$ on nende katseisikute arv, kellel toimub ajahetkel t huvialune sündmus.

Peamiselt erinevad Coxi põhjusjärgse riski mudel ja Fine-Gray alamjaotuse riski mudel selle poolest, kuidas neis kasutatakse konkureerivaid sündmusi ning kuidas saab mudelikordajaid tõlgendada. Fine-Gray mudelis käsitletakse konkureerivate sündmuste toimumist informatiivse tsenseerimisena, samas kui Coxi mudelis käsitletakse konkureeriva sündmuse toimumist tavalise tsenseerimisena. (Zhang, 2016) Seega jäetakse Coxi mudelis katsealused, kellel on toimunud konkureeriv sündmus enne huvialust sündmust, riskigrupist välja ning Fine-Gray mudelis jäetakse nad riskigruppi alles. Vaatamata sellele, et Fine-Gray mudelis käsitletakse konkureeriva sündmuse toimumist informatiivse tsenseerimisena, ei hinda Fine-Gray alamriski mudel riskifunktsiooni paremini kui Coxi põhjusjärgse riski mudel. (Allison, 2018)

Tunnuste mõju kumulatiivsele põhjusespetsiifilisele esinemusfunktsioonile Coxi mudeliga otse hinnata ei saa, kuid seda saab teha Fine-Gray mudeli abil. See-eest tunnuste mõju hindamiseks riskifunktsioonile soovitatakse Coxi mudelit, kuna see on enamasti täpsem kui Fine-Gray mudel. Seega on üldine soovitus, et kui eesmärk on hinnata tunnuste mõju kumulatiivsele esinemusfunktsioonile, tuleks kasutada Fine-Gray mudelit, kuid kui eesmärk on hinnata tunnuste mõju riskile, tuleks kasutada Coxi põhjusjärgse riski mudelit. (Collett, 2015; Allison, 2018; Austin ja Fine, 2017) Siiski leidub ka autoreid, kes Fine-Gray mudeli kasutamist pigem ei soovita (Allison, 2018; Therneau, Crowson ja Atkinson, 2024).

1.3 Coxi ja Fine-Gray mudeli võrdlus simulatsiooni abil

Võrdlemaks Coxi mitme seisundiga võrdeliste riskide mudeli ning Fine-Gray võrdeliste alamriskide mudeli hinnanguid illustreerivalt, viidi läbi simulatsiooniuring. Genereeriti valim suurusega 10 000, milles olid järgmised tunnused:

1. Bernoulli jaotusega grupeeriv tunnus

$$X \sim Be(0,5).$$

2. Weibulli jaotusega elukestus sünnist surmani:

$$T_S \sim Weib(\lambda_S, \gamma_S),$$

kus $\gamma_S = 6$ ning $\lambda_S = e^{4.35 - b_1 \cdot X^1}$, kus $b_1 = \frac{\ln(1,5)}{6}$, millest järeldub, et kahe grupi riskifunktsioonide suhe on 1,5.

3. Weibulli jaotusega aeg diabeedini:

$$T_D \sim Weib(\lambda_D, \gamma_D),$$

kus $\gamma_D = 6$ ning $\lambda_D = e^{4.5 - b_2 \cdot X}$ ja $b_2 = \frac{\ln(2)}{6}$ (vastav riskide suhe on 2).

4. Ühtlase jaotusega aeg tsenseerimiseni:

$$Z \sim U(20; 105).$$

Simuleeritud andmetel defineeriti elukestus kui vähim aegadest diabeedi, surma või tsenseerimiseni ning sündmuse indikaator, mis võttis väärtuse 0, kui vähim oli aeg tsenseerimiseni, väärtuse 1, kui vähim oli aeg teise tüübi diabeedi saamiseni

¹Siin kasutame Weibulli jaotuse parametriseringut, mille korral elulemusfunktsioon avaldub kui $S(t) = \exp(-(\lambda t)^\gamma)$

ning väärtuse 2, kui vähim oli aeg surmani. Seejärel hinnati Coxi ja Fine-Gray mudelite abil teise tüübi diabeedi diagnoosi saamise ja suremise tõenäosused enne 70-aastaseks saamist. Mudelite abil hinnati need tõenäosused tunnuse X põhjal moodustatud gruppides. Lisaks leiti tunnuse X moodustatud gruppides erinevad simuleeritud andmete tegelikud tõenäosused. Tõenäosused leiti klassikalise tõenäosuse põhimõttel, eeldades et kõigi isikute ajad T_S ja T_D on teada. Andmete genereerimist, mudelite hindamist ning tõenäosuste arvutamist korrati kokku 100 korda.

Tähistame Coxi mudeli abil hinnatud tõenäosused kui \hat{P}_{Cox} ning Fine-Gray mudeli abil hinnatud tõenäosused kui \hat{P}_{FG} . Tähistame simuleeritud andmetelt arvatud tõenäosused järgnevalt:

1. P_1 - tõenäosus, et vaadeldav sündmus toimub enne vanust 70 ja enne konkureeriva sündmuse toimumist;
2. P_2 - tõenäosus, et vaadeldav sündmus toimub enne vanust 70, kui konkureerivat sündmust ei saa toimuda.

Tabelis 1 on välja toodud keskmised mudelite hinnangud haigestumise ja suremise tõenäosustele ning keskmised simuleeritud andmete tegelikud tõenäosused P_1 ja P_2 . Tabelist tundub, et mõlemad mudelid hindavad tõenäosust P_1 ehk tõenäosust, et vaadeldav sündmus toimub enne konkureeriva sündmuse toimumist ja enne vanust 70. Lisaks on tabelis 1 toodud mudelite kordajate eksponendid. Nähtub, et Coxi mudel hindab tegelikke riskide suhteid 1,5 ja 2 üsna hästi, kuid Fine-Gray mudelid neid ei hinda.

Tabel 1: Simulatsiooni käigus saadud hinnangute ja tõenäosuste keskmised.

| S | X | \hat{P}_{Cox} | \hat{P}_{FG} | P_1 | P_2 | $\exp(\hat{\beta}_{Cox})$ | $\exp(\hat{\beta}_{FG})$ |
|------|---|-----------------|----------------|-------|-------|---------------------------|--------------------------|
| T2D | 0 | 0,154 | 0,167 | 0,154 | 0,198 | | |
| surm | 0 | 0,379 | 0,403 | 0,380 | 0,419 | | |
| T2D | 1 | 0,251 | 0,239 | 0,251 | 0,358 | 2,003 | 1,501 |
| surm | 1 | 0,464 | 0,441 | 0,464 | 0,558 | 1,502 | 1,129 |

Tabelis 2 on välja toodud mudelihinnangute keskmised ruuterinevused tegelike tõenäosuste ning riskide suhetega. Ka sellest tabelist nähtub, et mõlemad mudelid hindavad üpris täpselt tõenäosust, et vaadeldav sündmus toimub enne uuritavat ajapunkti ja enne konkureeriva sündmuse toimumist. Siiski on ka siit tabelist näha, et Coxi mudel hindab tõenäosust paremini. Samuti on tabelist 2 näha, et tegelikke riskide suhteid hindab paremini Coxi mudel.

Tabel 2: Simulatsiooni käigus saadud hinnangute ja tõenäosuste keskmised ruuterinevused.

| S | X | P_1, \hat{P}_{Cox} | P_1, \hat{P}_{FG} | P_2, \hat{P}_{Cox} | P_2, \hat{P}_{FG} | $\exp(\beta_{Cox})$ | $\exp(\beta_{FG})$ |
|------|---|----------------------|---------------------|----------------------|---------------------|---------------------|--------------------|
| T2D | 0 | 3e-05 | 0,00019 | 0,00194 | 0,00099 | | |
| surm | 0 | 4e-05 | 0,00059 | 0,00162 | 0,00030 | | |
| T2D | 1 | 3e-05 | 0,00018 | 0,01147 | 0,01409 | 0,01218 | 0,25449 |
| surm | 1 | 5e-05 | 0,00058 | 0,00888 | 0,01357 | 0,00299 | 0,13927 |

1.4 Tõkestatud keskmise eluea hinnang

Kahes eelnenud alapeatükis kirjeldati kahte olulist võrdeliste riskide eeldusega meetodit haiguse põhjusespetsiifilise absoluutriski hindamiseks konkureerivate riskide korral. Alati ei pruugi aga võrdeliste riskide eeldus täidetud olla. Käesolevas alapeatükis on kirjeldatud elulemusanalüüsi meetodit, mis ei põhine võrdeliste riskide eeldusel, ning mida saab kasutada ka siis, kui vaadeldav sündmus toimub väga harva (Han ja Jung, 2022).

Tõkestatud keskmise eluea hinnangu abil hinnatakse keskmist arvu aastaid haiguseni teatud ajaperioodi jooksul. Seega pakub meetod lihtsasti tõlgendatava alternatiivi teistele elulemusanalüüsi meetoditele gruppide riskide võrdlemiseks. Juhul, kui konkureerivaid riske ei esine, on tõkestatud keskmine eluiga t^* aasta jooksul defineeritud kui

$$e(0,t^*) = \int_0^{t^*} S(t)dt$$

ehk vaadeldakse elulemusfunktsiooni integraalialust pindala ajahetkeni t^* . (Han ja Jung, 2022; Andersen, 2013)

Konkureerivate riskide korral kasutatakse tõkestatud keskmise eluea hindamisel elulemusfunktsiooni asemel põhjusespetsiifilist kumulatiivset haigestumusfunktsiooni. Selle abil hinnatakse nii-öelda kaduma läinud aastate arv, mida saab haigusriskist rääkides tõlgendada kui keskmist aega haiguse diagnoosist vaadeldud ajahetkeni t^* . Tõkestatud keskmine põhjusjärgselt kaduma läinud aastate arv on seega defineeritud kui

$$L_j(0,t^*) = \int_0^{t^*} F_j(t)dt.$$

Seos nende kahe funktsiooni vahel näiteks juhul, kui konkureerivaid riske ei esine, on $e(0,t^*) = t^* - L(0,t^*)$. (Andersen, 2013; Mozumder, Rutherford ja Lambert, 2021)

Enamasti huvitatakse aga absoluutriski hinnangust teatud tunnuste korral ja seda ka siis kui võrdeliste riskide ega võrdeliste alamjaotuse riskide eeldus ei ole täidetud. Siis ei saa põhjusjärgse kumulatiivse haigestumuse hindamiseks käesolevas töös seni vaadeldud meetodeid kasutada ja seega ei saa ka põhjusjärgset keskmist kaduma läinud aastate arvu arvutada. Sellisel juhul kasutatakse tõkestatud keskmise eluea hindamisel pseudovaatluseid, mille abil tehakse regressioonianalüüs tunnuste ja keskmise eluea vaheliste seoste hindamiseks. (Parner, Andersen ja Overgaard, 2023; Andersen ja Pohar Perme, 2010)

Tähistame Aalen-Johanseni meetodi abil saadud hinnangu põhjusespetsiifilisele kumulatiivsele haigestumusfunktsioonile, mis on saadud i -nda katseisiku valimist välja jätmise korral, kui $\hat{F}_j^{-(i)}(t^*)$. Siis selle katseisiku pseudovaatlus huvipakkuva sündmuse, mille korral $C = j$, jaoks ajahetkel t^* on defineeritud kui

$$\hat{\theta}_i(t^*) = n\hat{F}_j(t^*) - (n-1)\hat{F}_j^{-(i)}(t^*),$$

kus n on valimimaht. Sellisel moel saadud pseudovaatluseid nimetatakse ka *jack-knife* pseudovaatlusteks, kuid käesolevas töös on neid ka edaspidi lihtsalt pseudovaatlusteks nimetatud. Seejuures, kui konkureerivaid riske ei esine, kasutatakse pseudovaatluste genereerimisel Aalen-Johanseni hinnangute asemel Kaplan-Meieri hinnanguid. (Conner *et al.*, 2022; Andersen ja Pohar Perme, 2010)

Pseudovaatlustele hinnatakse üldistatud lineaarne mudel kui

$$g\left(E\left(\hat{\theta}_i(t^*)\right)\right) = \beta' \mathbf{x}_i.$$

Seejuures g tähistab mudeli seosefunktsiooni. Hinnang põhjusjärgsele kumulatiivsele haigestumusele saadakse avaldisele seosefunktsiooni pöördfunktsiooni rakendades. Mudelikordajad β hinnatakse üldistatud hinnanguvõrrandisüsteemi abil. Tähistagu V_i dispersioonifunktsiooni. Siis üldistatud hinnanguvõrrandisüsteem on

$$U(\beta) = \sum_{i=1}^n \left(\frac{\partial}{\partial \beta} g^{-1}(\beta' \mathbf{x}_i) V_i^{-1} (\hat{\theta}_i - g^{-1}(\beta' \mathbf{x}_i)) \right) = 0.$$

Hinnang parameetervektorile β saadakse võrrandi lahendamisel. (Conner *et al.*, 2022; Andersen, Hansen ja Klein, 2004)

Analoogiliselt pseudovaatluste abil kumulatiivse põhjusjärgse haigestumuse hindamisele saab pseudovaatluste abil hinnata ka põhjusjärgselt kaduma läinud aastate arvu. Tähistame $\hat{L}_j^{(-i)}$ hinnangu j -nda põhjuse järgselt kaduma läinud aastate arvule, mis on saadud i -nda katseisiku valmist välja jätmisel Aalen-Johanseni hin-

nangu abil. Siis i -s pseudovaatlus põhjusjärgsele kaduma läinud aastate arvule on defineeritud kui

$$J_j^i = n\hat{L}_j(t^*) - (n-1)\hat{L}_j^{-(i)}(t^*),$$

kus $\hat{L}_j(t^*)$ on hinnang kaduma läinud aastate arvule kogu valimis ning n valimimaht. Ka pseudovaatluste abil kaduma läinud aastate arvu hindamiseks hinnatakse pseudovaatlustele üldistatud lineaarne mudel, mille parameetrid hinnatakse üldistatud hinnanguvõrrandite abil. (Andersen, 2013)

Üldistatud lineaarse mudeli kordajad väljendavad muutust vastavalt põhjusjärgses kumulatiivses haigestumuses või põhjusjärgses kaduma läinud aastate arvus, millele on rakendatud seosefunktsiooni. Seega, kui kasutatakse lineaarset mudelit kaduma läinud aastate hindamiseks, saab mudelikordajaid tõlgendada muutuse-na kaduma läinud aastate arvus tunnuse muutudes ühe ühiku võrra ehk sisuliselt näitavad need, kui palju pikendab mõni tunnus keskmist aega huvialuse sündmuse toimumisest vaadeldava hetkeni. Analoogiliselt saab neid kordajaid tõlgendada sellena, kui palju lühemaks jääb tervena elatud aastate arv vastava tunnuse muutudes ühe ühiku võrra. Kasutades näiteks kumulatiivse haigestumuse hindamise mudelis logaritmilist seosefunktsiooni, saab mudelikordajaid tõlgendada logaritmilise riskide suhtena. (Conner ja Trinquart, 2021) Kasutades pseudovaatluste abil kumulatiivse haigestumuse hindamisel täiendavat log-log seosefunktsiooni, peaksid mudelikordajad sarnanema Fine-Gray mudeli omadega (Therneau, 2024a).

Praktikas kasutatakse analüüsis vahel *jack-knife* pseudovaatluste asemel lõpmata väikse *jack-knife* meetodil saadud pseudovaatluseid. Näiteks tehakse seda paketi *survival* funktsioonis *pseudo*. Sel meetodil saadud pseudovaatlused põhinevad esimese astme Taylori rittaarendusel ning nende peamine eelis tavapärase *jack-knife* pseudovaatluste ees on arvutuslik kiirus. Mõlemal meetodil saadud hinnangud on omavahel väga sarnased. (Therneau, 2024a; Therneau, 2024b)

2 60—90-aastaste geenidoonorite teise tüüpi diabeedi riski hinnangud

Käesoleva magistritöö teine peatükk on jagatud kolmeks alapeatükiks. Neist esimeses alapeatükis on kirjeldatud töös kasutatud andmeid ning nende päritolu. Peatüki teises alapeatükis on kirjeldatud, kuidas esimeses alapeatükis kirjeldatud andmeid töö käigus analüüsitud on. Kolmandas alapeatükis on kokku võetud tähtsamad tulemused, mis töö käigus saadud on.

2.1 Ülevaade andmetest

Käesolevas magistritöös on kasutatud andmeid Tartu Ülikooli Eesti geenivaramu loodud riiklikust biopangast. Andmete kogumine algas aastal 2002, millest saati on TÜ Eesti geenivaramu biopangaga liitunud ligikaudu 210 000 Eesti täisealist elanikku. Seejuures enamik geenidoonoritest liitus biopangaga aastatel 2018 ning 2019. Kõigi geenidoonorite kohta on geenivaramusse kogutud nende terviseandmeid ning tehtud on ka pärilikkusaine analüüs. Seega võimaldavad biopangast saadud andmed analüüsida seoseid geeni- ning terviseandmete vahel. ([Tartu Ülikooli Eesti geenivaramu 2025](#))

Töös on kasutatud nende inimeste terviseandmeid, kes on olnud geenivaramuga liitudes enam kui 60, kuid kõige rohkem 90 aastat vanad. Mudelites on kasutatud nende geenivaramuga liitumise vanust, sugu, pikkust (cm), kaalu (kg), kehamassiindeksit (*KMI*) ja mõõdetud glükoositaset (mmol/L). Samuti on mudelites kasutatud indikaatortunnuseid selle kohta, kas inimene suitsetab (*suits*) ning kas tal on diagnoositud südame isheemiatõbi (*CAD*), kõrgvererõhktõbi (*HT*) või müokardiinfarkt (*MI*). Lisaks geenidoonorite terviseandmetele on töös riski hindamisel olulise tunnusena kasutatud teise tüüpi diabeedi standardiseeritud polügeenset riskiskoori (*PRS*).

Analüüsist on välja jäetud 1277 inimest, kes liitusid geenivaramuga pärast aastat 2021. Seda seetõttu, et viimane teist tüüpi diagnoosi saamise kuupäev kasutatud andmestikus oli 2020. aasta lõpus. Samuti on analüüsist välja jäetud 5994 geenidoonorit, kellel on teise tüüpi diabeet diagnoositud enne geenivaramuga liitumist. Lisaks on analüüsist välja jäänud inimesed, kelle kohta ei ole mõne eelnenud lõigus mainitud tunnuse väärtus teada ja ka need geenidoonorid, kelle kehamassiindeks on andmestikus väiksem kui 15 või suurem kui 50, kuna need on võimalikud sisestusvead.

Pärast eelnevalt kirjeldatud elimineerimist on analüüsi jäänud 28 450 geenidoonorit. Neist geenidoonoritest 1322 on diagnoositud 10 aasta jooksul geenivaramuga liitumisest teise tüüpi diabeet. 10 aasta jooksul alates geenivaramuga liitumise kuupäevast on surnud analüüsi jäänud geenidoonoritest 3209 geenidoonorit. Seejuures ei võeta arvesse neid geenidoonoreid, kellel diagnoositi enne surma teise tüüpi diabeet. Kasutatud tunnuste kirjeldavad statistikud on toodud tabelis 3.

Tabel 3: Kasutatud tunnuste kirjeldavad statistikud soo lõikes. Pidevate tunnuste puhul on toodud keskmine (standardhälve), binaarsete tunnuste puhul arv (%).

| | Mehed n=8846 (31,1%) | Naised n=19604 (68,9%) |
|------------------------------|-------------------------|---------------------------|
| vanus liitudes (aastates) | 69 (6,39) | 69 (6,35) |
| kaal (kg) | 85 (14,19) | 73 (13,62) |
| pikkus (cm) | 176 (6,67) | 162 (5,92) |
| kehamassiindeks (KMI) | 27,37 (4,05) | 27,74 (4,93) |
| vereglükoosi tase (mmol/L) | 4,40 (1,40) | 4,28 (1,26) |
| praegune suitsetaja | 1368 (15,5%) | 1656 (8,4%) |
| Liitumishetkeks diagnoositud | | |
| müokardiinfarkt (MI) | 2693 (30,4%) | 4144 (21,1%) |
| südame isheemiatõbi (CAD) | 2756 (31,2%) | 4397 (22,4%) |
| kõrgvererõhktõbi (HT) | 5872 (66,4%) | 12511 (63,8%) |
| Jälgimisajal vaadeldud | | |
| teist tüüpi diabeet (T2D) | 446 (5,0%) | 876 (4,5%) |
| surm | 1516 (17,1%) | 1693 (8,6%) |

2.2 Analüüsi käik

Andmete analüüsimisel hinnati Coxi mitme seisundiga mudel, Fine-Gray mudel ning üldistatud lineaarne mudel, mis hinnati pseudovaatluste abil tõkestatud keskmise eluea meetodil. Kõigi mudelite abil hinnati tõenäosust saada 10 aasta jooksul teist tüüpi diabeedi diagnoos. Lisaks hinnati pseudovaatluste abil 10 aastaga tõkestatud keskmist teist tüüpi diabeediga kaduma läinud aastate arvu.

Mudelite hindamiseks loodi kolme tasemega faktortunnus, mille 0-tase väljendas tsenseerimist, esimene tase, et doonoril diagnoositi 10 aasta jooksul teist tüüpi diabeet ning teine tase, et doonor suri kümne aasta jooksul alates geenivaramuga liitumisest. Tunnus võttis väärtuse 1, kui teist tüüpi diabeedi diagnoos saadi enne surma ning väärtuse 2, kui doonor suri mõnel muul põhjusel enne teist tüüpi diabeedi diagnoosi saamist. Doonoritel, kellel kumbagi neist sündmustest ei juhtunud, võttis tunnus väärtuse 0 ehk nad tsenseeriti. Lisaks loodi pidev tunnus, mis väljendas aega geenivaramuga liitumisest kirjeldatud sündmuse toimumiseni aastates. Tsenseeritud tunnuste puhul oli tsenseerimiskuupäevaks 31.12.2021, et tagada vähemalt aastane jälgimisperiood kõigile geenidoonoritele, kelle andmeid analüüsis kasutati.

Need geenidoonorid, kelle aeg geenivaramuga liitumisest sündmuse toimumiseni ületas 10 aastat, tsenseeriti 10. aastal. See tähendab, et nende sündmust indikeeriv tunnus võttis väärtuse 0 ning aega mõõtev tunnus väärtuse 10. Aega lõigati ning vaatlused tsenseeriti selleks, et ajaskaala sündmusteni ei oleks liiga pikk ning 10 aasta haigestumise riski oleks võimalik paremini hinnata.

Mudelite hindamisel kasutati tunnuste mudelisse valimisel segavalikut. Coxi konkureerivaid riske arvesse võtva võrdeliste riskide mudeli hindamiseks kasutati paketi *riskRegression* funktsiooni *CSC*, mis võimaldab kasutada erinevate põhjuste spetsiifiliste riskifunktsioonide hindamisel erinevaid tunnuseid. Paketi *survival* funktsioon *coxph* võimaldab küll samuti hinnata konkureerivaid riske arvesse võtvaid

riskifunktsioone, kuid seda kasutades peavad erinevate põhjuste järgsetes riskifunktsioonides olema samad tunnused. Võrdluseks hinnati Coxi mudelid ka *coxph* funktsiooniga. Ootuspäraselt selgus, et erinevaid funktsioone kasutades, jäid Coxi mudelites olulisteks samad tunnused ja nende koosmõjud. Seetõttu on magistritöö järgmises alapeatükis kirjeldatud vaid funktsiooni *CSC* abil hinnatud mudelit.

Fine-Gray võrdeliste alamriskide mudel hinnati paketi *riskRegression* funktsiooniga *FGR*. Nagu Coxi mudelitki, saab ka Fine-Gray mudelit hinnata ka paketi *survival* funktsiooniga. Selleks on seal funktsioon *finegray*. Mõlema paketi puhul võttis Fine ja Gray võrdeliste alamriskide mudeli hindamine Coxi võrdeliste riskide mudeli hindamisest rohkem aega.

Tõkestatud keskmise eluea analüüsi tarbeks loodi pseudovaatlued paketi *survival* funktsiooniga *pseudo*. Nende abil hinnati üldistatud lineaarne mudel funktsiooniga *glm*. Käesolevas analüüsis kasutati pseudovaatluseid, et hinnata tõenäosust, et inimesel on kümne aasta pärast teise tüüpi diabeedi diagnoos. Lisaks arvutati pseudovaatlused, millega hinnata eelnevates mudelites kasutatud tunnuste mõju keskmisele aastate arvule, mis 10st aastast diabeedis veedetakse.

Hinnatud mudeleid võrreldi omavahel peamiselt riskigruppe võrreldes ning jooniste abil. Esmalt uuriti, milline, on kolme meetodi kooskõla indiviidide rühmitamisel madala, keskmise ja kõrge riskiga rühmadesse. Seejärel vaadeldi, kas mudelite abil hinnatud erineva riskitasemega gruppides on ka Aalen-Johanseni funktsiooni abil hinnatud elulemuskõverad oluliselt erinevad ning millise mudeli abil hinnatud riskid on kõige sarnasemad Aalen-Johanseni hinnangutega.

Võrdluste põhjal valiti mudel, mille abil saaks kõige paremini anda geenidoonoritele tagasisidet nende T2D 10 aasta riski kohta. Lõpuks loodi välja valitud mudeli abil hinnatud riske illustreeriv joonis selle mudeli prognoositud riskidega mõnele hüpoteetilisele isikule.

2.3 Tulemused

Töö käigus hinnati teise tüüpi diabeeti (T2D) haigestumise riski kolme erineva mudeliga, mis võtsid kõik arvesse surma konkureeriva riskina. Coxi põhjusjärgse võrdeliste riskide mudeli põhjal mõjutavad teist tüüpi diabeedi saamist selle polügeenne riskiskoor (PRS) ning inimese kaal, pikkus, kehamassiindeks (KMI), glükoosinäit ja indikaatortunnus, kas inimesel on diagnoositud kõrgvererõhktõbi (HT). Lisaks mõjutab mudeli põhjal teise tüüpi diabeedi saamist kaalu ja kehamassiindeksi koosmõju. Tabelis 4 on iga tunnuse kohta välja toodud selle kordaja väärtus mudelis, kordaja standardviga ning tunnuse olulisuse tõenäosus mudelis. Lisaks on välja toodud samad näitajad suremuse riskimudeli kohta, kuid käesolevas magistritöös keskendutakse teise tüüpi diabeeti haigestumise riskile. Tunnus *suits* suremise riskimudelis on indikaatortunnus, mis võtab väärtuse 1, kui inimene suitsetab. Tunnus *sugu* on indikaator, kas tegemist on naisega ehk tunnuse väärtus on 1, kui geenidoonor on naine ning 0, kui doonor on mees. Tabelis 4 toodud mudeli hindamiseks kulus 0,36 sekundit.

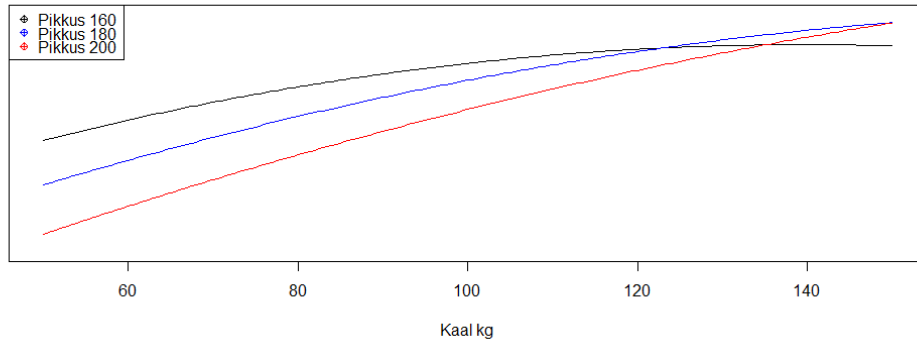
Tabel 4: Coxi võrdeliste riskide mudeli parameetrite hinnangud.

| | $\hat{\beta}$ | Standardviga | p-väärtus |
|--------------|---------------|--------------|-----------|
| T2D mudel | | | |
| PRS | 0,461 | 0,028 | <0,001 |
| KMI | -0,066 | 0,049 | 0,179 |
| kaal | 0,124 | 0,024 | <0,001 |
| pikkus | -0,089 | 0,02 | <0,001 |
| glükoos | 0,357 | 0,013 | <0,001 |
| HT | 0,432 | 0,072 | <0,001 |
| KMI: kaal | -0,001 | <0,001 | <0,001 |
| Surma mudel | | | |
| vanus | -0,094 | 0,066 | 0,155 |
| KMI | -0,027 | 0,033 | 0,422 |
| kaal | -0,102 | 0,014 | <0,001 |
| pikkus | -0,031 | 0,032 | 0,335 |
| sugu | -3,982 | 0,525 | <0,001 |
| HT | 0,264 | 0,043 | <0,001 |
| suits | 2,586 | 0,541 | <0,001 |
| KMI*kaal | 0,002 | <0,001 | <0,001 |
| vanus*sugu | 0,047 | 0,007 | <0,001 |
| vanus*suits | -0,026 | 0,008 | 0,001 |
| vanus*pikkus | 0,001 | <0,001 | 0,003 |

Paneme tähele, et pikkuse ja kaalu mõjud mudeli lineaarses osas avalduvad kui

$$0,124 \cdot \text{kaal} - 0,089 \cdot \text{pikkus} - 0,0066 \cdot \text{KMI} - 0,001 \cdot \text{KMI} \cdot \text{kaal},$$

kus $KMI = \frac{\text{kaal}}{(\text{pikkus}/100)^2}$. Mudeli paremaks tõlgendamiseks loodi joonis, kus y-teljel on saadud avaldise väärtus ja x-teljel kaal (joonis 1). Jooniselt 1 on näha, et suurema kehakaaluga 60 – 90-aastastel inimestel on suurem risk haigestuda teise tüüpi diabeeti, kusjuures sama suure kaalu juures on risk suurem lühemal inimesel. Mõju on natukene teistsugune väga kõrge kaalu puhul. Jooniselt nähtub, et väga kõrge kehamassiindeksi puhul jääb mõju riskile ühtlaselt kõrgeks ka kehamassiindeksi kasvades.



Joonis 1: Pikkuse ja kaalu mõju teise tüüpi diabeedi riskile.

Fine-Gray võrdeliste alamriskide mudeli põhjal mõjutavad kümne aasta jooksul teist tüüpi diabeedi diagnoosi saamist teise tüüpi diabeedi polügeenne riskiskoor ning inimese pikkus, kaal, kehamassiindeks glükoosinäit ja indikaator, kas tal on diagnoositud kõrgvererõhktõbi. Samuti mõjutab Fine-Gray mudeli põhjal teise tüüpi diabeedi saamist kaalu ja kehamassiindeksi koosmõju. Seega kasutavad Coxi ja Fine-Gray mudel diabeedispetsiifilise riski hindamisel samu tunnuseid. Tunnuste kordajate hinnanguid vastavate standardvigade ja p-väärtustega saab näha tabelist 5. Tabelis kirjeldatud mudeli hindamiseks kulus 42,87 sekundit.

Tabel 5: Fine-Gray võrdeliste alamriskide mudeli parameetrite hinnangud.

| | $\hat{\beta}$ | Standardviga | p-väärtus |
|----------|---------------|--------------|-----------|
| PRS | 0,465 | 0,029 | <0,001 |
| KMI | -0,060 | 0,051 | 0,24 |
| kaal | 0,132 | 0,024 | <0,001 |
| pikkus | -0,094 | 0,021 | <0,001 |
| glükoos | 0,351 | 0,016 | <0,001 |
| HT | 0,401 | 0,074 | <0,001 |
| KMI*kaal | -0,001 | <0,001 | <0,001 |

Tabelitest 5 ja 4 on näha, et tunnuste mudelikordajate ning standardvigade väärtused on Coxi võrdeliste riskide ning Fine-Gray võrdeliste alamriskide teist tüüpi

diabeedi riskimudelites üsna sarnased. Mudelikordajate suunad on ootuspärased. Mida suurem on inimese polügeenne riskiskoor, seda suurem on tema risk haigestuda teist tüüpi diabeeti. Samuti on risk teise tüüpi diabeeti haigestuda seda kõrgem, mida kõrgem on inimese keskmine glükoosinäit. Lisaks on teistest kõrgem risk haigestuda teise tüüpi diabeeti neil 60 – 90-aastastel inimestel, kellel on diagnoositud kõrgvererõhktõbi.

Lisaks Coxi ja Fine-Gray mudelitele, hinnati 10 aasta haigestumise riski ka pseudovaatluste abil. Mudeliga hinnati tõenäosust, et inimesel on 10 aasta pärast teise tüüpi diabeet. Selleks, et mudel oleks võrreldav Coxi ning Fine ja Gray mudelitega, kasutati mudelis esmalt samu tunnuseid ja nende koosmõjusid, mida kahes eelnenud mudelis diabeedi riski hindamisel kasutati. Hinnanguid mudelikordajatele nende standardvigade ja p-väärtustega saab näha tabelist 6, kus on välja toodud ka mudeli vabaliige. Tagamaks, et mudelikordajad oleksid võrreldavad Coxi ja Fine-Gray mudelite kordajatega, kasutati mudeli hindamisel täiendavat log-log linki.

Tabel 6: Pseudovaatlustega mudel tõenäosusele.

| | $\hat{\beta}$ | Standardviga | p-väärtus |
|-------------|---------------|--------------|-----------|
| (Vabaliige) | 5,834 | 2,981 | 0,05 |
| PRS | 0,463 | 0,028 | <0,001 |
| KMI | -0,072 | 0,045 | 0,108 |
| kaal | 0,134 | 0,023 | <0,001 |
| pikkus | -0,101 | 0,019 | <0,001 |
| glükoos | 0,429 | 0,016 | <0,001 |
| HT | 0,436 | 0,076 | <0,001 |
| KMI*kaal | -0,001 | <0,001 | <0,001 |

Ootuspäraselt on näha, et mudelikordajate ning nende standardvigade hinnangud sarnanevad Coxi ning Fine ja Gray T2D riskimudelite kordajate ja nende standardvigade kordajatega. Samuti on näha, et kõik tunnused on mudelis statistiliselt olulised. Tabelis kirjeldatud mudeli hindamiseks kulus 0,12 sekundit. Mudelis kasutatud pseudovaatluste arvutamiseks kulus ligikaudu 4,91 sekundit. Nägemaks,

kuidas mõjutavad tunnused keskmist aastate arvu, mis kümnest aastast diabeediga veedetakse, hinnati pseudovaatluste abil ka tõkestatud keskmise kaduma läinud aastate arvu mudel teise tüüpi diabeedile. Mudeli hindamisel kasutati identsusseost, et mudeli kordajad oleksid lihtsasti tõlgendatavad. Hinnatud mudelikordajaid nende standardvigade ja p-väärtustega saab näha tabelist 7. Tabelis kirjeldatud mudeli hindamiseks kulus 0,02 sekundit. Mudelis kasutatud pseudovaatluste arvutamiseks kulus ligikaudu 5,72 sekundit.

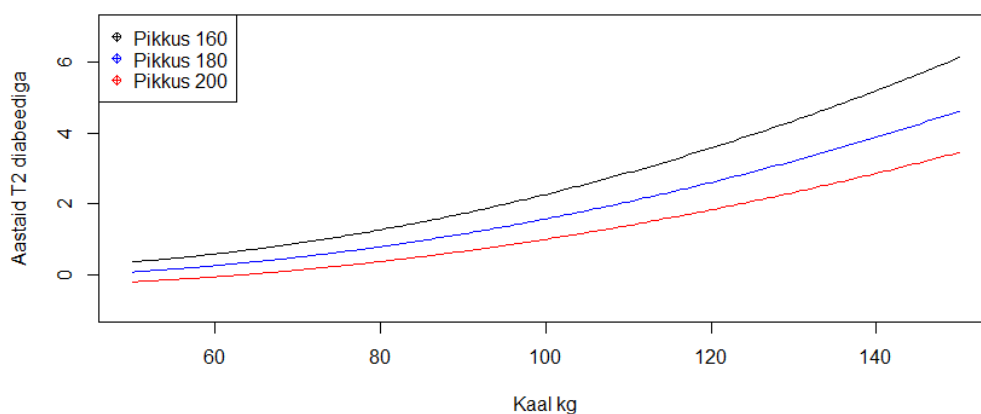
Tabel 7: Pseudovaatlustega mudel kaotatud aastate arvule.

| | $\hat{\beta}$ | Standardviga | p-väärtus |
|-------------|---------------|--------------|-----------|
| (Vabaliige) | 1,694 | 1,523 | 0,266 |
| PRS | 0,207 | 0,014 | <0,001 |
| KMI | -0,050 | 0,027 | 0,064 |
| kaal | -0,001 | 0,012 | 0,945 |
| pikkus | -0,014 | 0,01 | 0,15 |
| glükoos | 0,240 | 0,01 | <0,001 |
| HT | 0,075 | 0,028 | 0,008 |
| KMI*kaal | 0,001 | <0,001 | <0,001 |

Tabelist 7 on näha, et keskmiselt saadakse T2D diagnoos umbes 1,7 aastat enne 10 aasta möödumist alguspunktist. Seejuures glükoosinäidu tõus ühiku võrra põhjustab keskmiselt veerand aastat varasema haigestumise. Polügeenne riskiskoor on analüüsis standardiseeritud, mistõttu saab selle kordajat mudelis tõlgendada nii, et need, kellel on keskmisest ühe standardhälbe võrra kõrgem polügeenne riskiskoor haigestuvad teist tüüpi diabeeti keskmiselt ligi kaks ja pool kuud varem kui keskmise T2D polügeense riskiskooriga inimesed. Mudeli põhjal saavad inimesed, kellel on diagnoositud kõrgvererõhktõbi, teist tüüpi diabeedi diagnoosi keskmiselt ligi kuu aega varem kui need, kellel kõrgvererõhktõbe ei ole.

Seda, kui palju muudavad keskmiselt 10 aasta jooksul diabeediga elatud aastaid kaalu, pikkuse ja kehamassiindeksi muutused, on otse mudelist üsna keeruline välja lugeda, kuna kehamassiindeks on kaalu ja pikkuse funktsioon. Sarnaselt sellele,

kuidas visualiseeriti joonisel 6 pikkuse ja kaalu mõju Coxi mudeli abil hinnatud teist tüüpi diabeedi riskile, illustreeriti mainitud tunnuste mõju keskmisele teist tüüpi diabeediga elatud aastate arvule 10st. Joonisel 2 on visualiseeritud kaalu muutuse mõju elada jäänud aastate arvule samade pikkuste korral, millega näitlikustati kaalu mõju Coxi põhjusjärgse mudeli puhul teise tüüpi diabeedile. Joonisel on võrreldud inimesi, kellel ei ole kõrgvererõhktõbe ning kelle standardiseeritud polügeense riskiskoori väärtus on 0 ja glükoosinäit 4.



Joonis 2: Pikkuse ja kaalu mõju teist tüüpi diabeediga elatud aastatele 10st.

Mudelite omavahel võrdlemiseks jaotati inimesed mudelite abil hinnatud riskide põhjal gruppidesse. Selleks hinnati iga mudeli abil inimese riski haigestuda 10 aasta jooksul teist tüüpi diabeeti. Saadud hinnangu põhjal liigitati inimene madala riskiga gruppi, kui tema risk oli alla 10%, keskmise riskiga gruppi, kui tema risk oli vahemikus 10—20% ning kõrge riskiga gruppi, kui tema mudeli abil hinnatud risk oli kõrgem kui 20%. Mudelite võrdlemiseks võrreldi, kui paljud inimesed on erinevate mudelite abil hinnatud riskide põhjal erinevates riskigruppides. Tabelis 8 on välja toodud inimeste arvud erinevates riskigruppides.

Tabel 8: Mudelite abil hinnatud riskigruppide kokkulangevused.

| | | Pseudovaatlused | | |
|----------|-----------|-----------------|-------------|-------------|
| | | madal | keskmine | kõrge |
| Cox | Fine-Gray | | | |
| madal | madal | 20735 | 1081 | 0 |
| madal | keskmine | 5 | 319 | 9 |
| keskmine | madal | 24 | 275 | 0 |
| keskmine | keskmine | 12 | 3375 | 659 |
| keskmine | kõrge | 0 | 0 | 111 |
| kõrge | keskmine | 0 | 3 | 129 |
| kõrge | kõrge | 0 | 2 | 1711 |

Tabelist 8 on näha, et geenidoonoreid, kelle risk jäi kolme mudeli abil hinnatud riskide põhjal samasse riskigruppi, oli 25 821. Kokku osales analüüsis 28 450 geenidoonorit. Seega oli mudelite hinnatud riskigrupp kõigi kolme mudeli puhul sama umbes 90%-l analüüsis osalenutest. Seejuures on inimesi, kellel oli mõne mudeli abil hinnatud risk madal, kuid mõne teise mudeli abil hinnatud risk kõrge, kokku vaid 9. Neist inimestest kuuel lõppes jälgimisaeg surmaga ning vaid ühel diagnoositi T2D. Tabelist näib, et keskmiselt võib pseudovaatlustega loodud riskimudeli abil hinnatud haigestumise risk olla pisut kõrgem kui Coxi mudeli abil hinnatud haigestumise risk. Võrreldes Coxi ja Fine-Gray mudelite abil hinnatud riskigruppides olevate indiviidide arve, näib, et keskmiselt võib Fine-Gray mudel riski hinnata pigem natukene väiksemaks, kui Coxi mudel.

Nägemaks, milline mudelitest riskigruppe kõige paremini hindab, jagati inimesed gruppidesse kõigi mudelite moodustatud riskigruppide põhjal ehk moodustati grupid, mille suuruseid on näha tabelist 8. Seejuures vaadeldi vaid grupe, mille suurus oli vähemalt 100. Gruppides leiti Aalen-Johanseni hinnang elulemusele ning võrreldi T2D riski hinnangut 10. aastal. Tabelis 9 on välja toodud riskigrupid ning nende suurused (n). Samuti on toodud nende inimeste arv, kes 10. aastal riskigruppi kuulusid ja ka nende arv, kes siis teist tüüpi diabeeti haigestusid. Arvud on tähistatud vastavalt n_{risk} ja n_{T2D} . Lisaks on toodud Aalen-Johanseni meetodil saadud

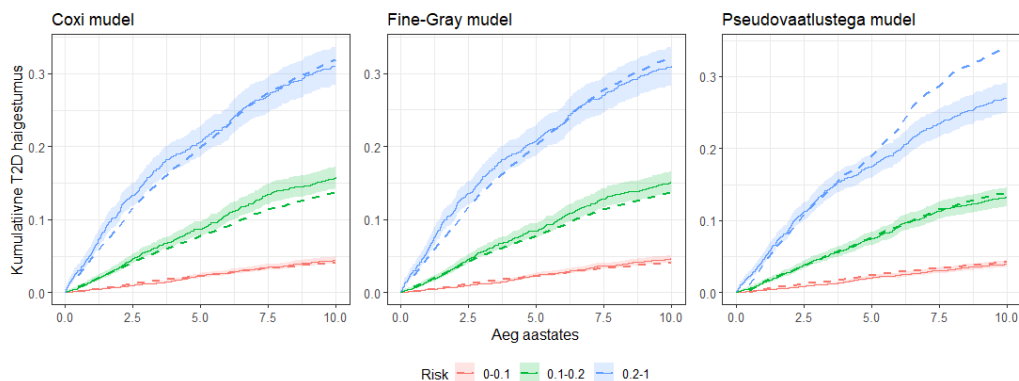
hinnang grupi keskmisele teist tüüpi diabeeti haigestumise tõenäosuselem mis on toodud veerus P_{T2D} . Tõenäosuse hinnang on tabelis toodud kümnendmurruna.

Tabel 9: Aalen-Johanseni hinnang T2D 10 aasta riskile riskigruppides.

| Cox | Fine-Gray | Pseudovaatlused | n | n_{risk} | n_{T2D} | P_{T2D} |
|-----------------|-----------------|-----------------|--------------|-------------|------------|--------------|
| madal | madal | madal | 20735 | 3367 | 369 | 0,040 |
| madal | madal | keskmine | 1081 | 295 | 66 | 0,098 |
| madal | keskmine | keskmine | 319 | 31 | 14 | 0,054 |
| keskmine | madal | keskmine | 275 | 89 | 23 | 0,147 |
| keskmine | keskmine | keskmine | 3375 | 830 | 308 | 0,152 |
| keskmine | keskmine | kõrge | 659 | 199 | 85 | 0,177 |
| keskmine | kõrge | kõrge | 111 | 14 | 16 | 0,180 |
| kõrge | keskmine | kõrge | 129 | 42 | 21 | 0,213 |
| kõrge | kõrge | kõrge | 1711 | 461 | 417 | 0,318 |

Tabelist 9 on näha, et kõige õigemini hindab riskigruppi Coxi mudel. Aalen-Johanseni meetodi abil saadud keskmise riski hinnang langeb antud tabeli puhul gruppides, kus eri meetodite hinnangud kokku ei langenud, alati samasse kategooriasse, milles on vastava grupi Coxi mudeliga hinnatud risk.

Lisaks võrreldi mudeleid ka visuaalselt, võrreldes mudelite abil hinnatud teist tüüpi diabeedi haigestumuskõveraid gruppides Aalen-Johanseni meetodiga hinnatud elulemuskõveraga. Haigestumuskõveraid võrdlevad joonised on joonisel 3, kus on pideva joonega Aalen-Johanseni hinnangud kumulatiivsele teist tüüpi diabeeti haigestumusele riskigruppides. Hinnangutele on toodud ka 95%-usaldusvahemikud. Lisaks on joonisel katkendliku joonega iga mudeli kohta vastava mudeli abil hinnatud kumulatiivne haigestumuskõver.



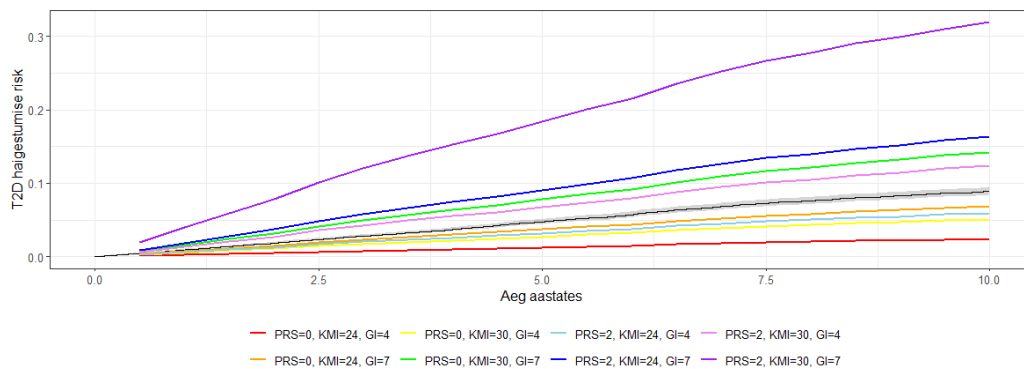
Joonis 3: Mudelite põhjal hinnatud haigestumuskõverate (punktirjoonega) võrdlused Aalen-Johanseni hinnangutega (pidevjoonega) kolmes erineva riskihinnangu grupis (vastavalt mudelipõhisele 10 aasta riski hinnangule).

Jooniselt 3 on näha, et pseudovaatluste abil hinnatud mudeli kõrge riskiga grupis on 10 aasta haigestumise risk keskmiselt liiga kõrgeks hinnatud. Coxi ja Fine-Gray mudelite abil hinnatud haigestumuskõverate puhul ei ole jooniselt nii suuri erinevusi Aalen-Johanseni hinnanguga näha. Lisaks on jooniselt näha, et Fine-Gray mudeli abil hinnatud risk on üsna sarnane Coxi mudeli abil hinnatud riskile.

Kolme mudeli võrdluse põhjal sobib teise tüüpi diabeedi 10 aasta haigestumise riski hindamiseks 60–90-aastastele eesti inimestele kõige paremini Coxi mitme seisundiga mudel. Analüüsist nähtus ka, et Fine-Gray mudel hindab haigestumise riski Coxi mudelile üsna sarnaselt, kuid Coxi mudeli abil jagatakse inimesed riskigruppidesse natukene täpsemalt. Samuti oli magistritöö esimeses peatükis kirjeldatud simulatsioonist näha, et Coxi mudel hindab põhjusjärgseid riske Fine-Gray mudelist täpsemalt. Fine-Gray mudelile tasub Coxi mudelit riski hindamisel eelistada ka seetõttu, et näiteks rakendustarkvaras R saab Coxi mudelit Fine-Gray mudelist märgatavalt kiiremini hinnata.

Nägemaks, kuidas erinevad Coxi mudeli abil hinnatud riskid erinevate polügeense riskiskooride, kehamassiindeksite ning glükoosinäitude korral, tehti joonis, millel on Coxi mudeli abil saadud teise tüüpi diabeedi haigestumise riski hinnanguid illustree-

ritud mõne hüpoteetilise isiku jaoks. Nendeks isikuteks olid 70-aastased naised kahe võimaliku kehamassiindeksiga ($KMI = 24$ ja $KMI = 30$), kahe võimaliku glükoositasemega ($GI = 4$ mmol/L ja $GI = 7$ mmol/L) ning kahe võimaliku polügeense riskiskoori väärtusega ($PRS = 0$ ja $PRS = 2$). Naiste pikkuseks võeti 162 cm ning kaal leiti vastavalt kehamassiindeksile. Nii saadi kokku kaheksa näidisinimest, kelle riskikõveraid saab näha joonisel 4. Lisaks on joonisel Aalen-Johanseni hinnang keskmisele riskikõverale, mille hindamisel kasutati tervet magistritöös kasutatud andmestikku.



Joonis 4: Hinnatud riskikõverate võrdlus 70-aastaste naiste seas.

Jooniselt 4 on näha, et võrreldes erinevate kehamassiindeksite, polügeensete riskiskooride ning glükoosinäitudega inimesi, mõjub iga tunnus teist tüüpi diabeeti haigestumise riskile nii, et kui kahe teise tunnused on võrreldavatel inimestel samad, on haigestumise risk kõrgem sellel, kellel on kõrgem vaadeldava tunnuse väärtus. Seejuures on joonisel vaadeldud gruppides kehamassiindeksist mõjukam glükoosinäit. Näiteks on risk kõrgem inimestel, kellel on keskmine geneetiline risk ja kehamassiindeks, kuid kõrge glükoosinäit kui neil inimestel, kellel on keskmine glükoosinäit ja PRS, kuid kõrge KMI.

Kokkuvõte

Käesoleva magistritöö eesmärk oli anda ülevaade konkureerivaid riske arvesse võtvatest haigusrisiki hindamise meetoditest ning võrrelda neist peamiseid Tartu Ülikooli Eesti geenivaramu andmetel teise tüüpi diabeedi riski hindamisel, et näha, millise mudeliga saaks kõige paremini hinnata vanemaealiste inimeste teist tüüpi diabeeti haigestumise riski. Töös võrreldi omavahel Coxi võrdeliste riskide mudelit, Fine-Gray võrdeliste alamriskide mudelit ning pseudovaatluste abil hinnatud lineaarset riskimudelit. Mudelite abil hinnati riski haigestuda teist tüüpi diabeeti, võttes seejuures arvesse võimalust enne haigestumist surra.

Coxi ning Fine ja Gray mudelite võrdlemiseks tehti simulatsioon, millest nähtus, et mõlemad mudelit hindavad samu tõenäosuseid, kuid Coxi mudeli abil saadud riskihinnangud on täpsemad. Töö praktilises osas võrreldi kolme eelmainitud mudeli riskihinnanguid omavahel ja ka Aalen-Johanseni hinnangutega. Lisaks mõõdeti, kui kaua võttis aega iga mudeli hindamine.

Analüüsist selgus, et kõige paremini sobib geenidoonorite teist tüüpi diabeeti haigestumise riski hindamiseks Coxi põhjusjärgne võrdeliste riskide mudel. Selle mudeliga saadud riskihinnangud olid kõige lähedasemad Aalen-Johanseni meetodil saadud hinnangutega. Lisaks võttis Coxi mudeli hindamine märgatavalt vähem aega kui Fine ja Gray mudeli hindamine.

Töös kasutatud Coxi mudel kasutas diabeedi riski hindamiseks standardiseeritud T2D polügeenset riskiskoori, inimese pikkust, kaalu, kehamassiindeksit, glükoosinäitu ja indikaatortunnust, kas inimesel on kõrgvererõhktõbi. Lisaks oli diabeedi riskimudelis KMI ja kaalu koosmõju. Coxi mudeli puhul kasutatakse põhjusjärgse absoluutrisiki hindamisel ka konkureeriva riski mudelit.

Töö alusel on võimalik täiendada Tartu Ülikooli Eesti geenivaramu tagasisideportaali MinuGeenivaramu, et anda detailsemat tagasisidet riskide osas ka eakamatele

inimestele. Selleks tuleks mudeliga veel tööd teha. Kaaluda tasuks ka võimalust hinnata diabeediriski parameetriliste mudelite abil, mida käesolevas magistritöös ei käsitletud.

Uuring oli teostatud väljastuse 3-10/GI/10790 raames ja saanud Eesti bioetika inimuuringu nõukogu kooskõlastuse (20.09.2023, nr 1.1-12/3455).

Kasutatud allikad

- Allison, P. (2018). For causal analysis of competing risks, don't use Fine & Gray's subdistribution method. *Statistical Horizons* 2022. URL: <https://statisticalhorizons.com/for-causal-analysis-of-competing-risks/> (vaadatud 23.02.2025).
- Andersen, P. K. ja Pohar Perme, M. (2010). Pseudo-observations in survival analysis. *Statistical methods in medical research* 19.1, lk. 71–99. DOI: <https://doi.org/10.1177/0962280209105020>.
- Andersen, P. K. (2013). Decomposition of number of life years lost according to causes of death. *Statistics in medicine* 32.30, lk. 5278–5285. DOI: <https://doi.org/10.1002/sim.5903>.
- Andersen, P. K., Hansen, M. G. ja Klein, J. P. (2004). Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime data analysis* 10, lk. 335–350. DOI: <https://doi.org/10.1007/s10985-004-4771-0>.
- Austin, P. C. ja Fine, J. P. (2017). Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Statistics in medicine* 36.27, lk. 4391–4400. DOI: <https://doi.org/10.1002/sim.7501>.
- Benichou, J. ja Gail, M. H. (1990). Estimates of Absolute Cause-Specific Risk in Cohort Studies. *Biometrics* 46.3, lk. 813–826. DOI: <https://doi.org/10.2307/2532098>.
- Collett, D. (2015). *Modelling Survival Data in Medical Research*. Bristol, UK: CRC Press LLC.
- Conner, S. C., Beiser, A., Benjamin, E. J., LaValley, M. P., Larson, M. G. ja Trinquart, L. (2022). A comparison of statistical methods to predict the residual lifetime risk. *European journal of epidemiology* 37 (2), lk. 173–194. DOI: <https://doi.org/10.1007/s10654-021-00815-8>.

- Conner, S. C. ja Trinquart, L. (2021). Estimation and modeling of the restricted mean time lost in the presence of competing risks. *Statistics in medicine* 40.9, lk. 2177–2196. DOI: <https://doi.org/10.1002/sim.8896>.
- Cox, D. ja Oakes, D. (2018). *Analysis of Survival Data*. Chapman and Hall/CRC.
- Fine, J. P. ja Gray, R. J. (1999). A proportional hazards model for the sub-distribution of a competing risk. *Journal of the American statistical association* 94.446, lk. 496–509. DOI: <https://doi.org/10.1080/01621459.1999.10474144>.
- Han, K. ja Jung, I. (2022). Restricted mean survival time for survival analysis: a quick guide for clinical researchers. *Korean Journal of Radiology* 23.5, lk. 495. DOI: <https://doi.org/10.3348/kjr.2022.0061>.
- Hansen, S. N., Overgaard, M., Andersen, P. K. ja Parner, E. T. (2017). Estimating a population cumulative incidence under calendar time trends. *BMC medical research methodology* 17. DOI: <https://doi.org/10.1186/s12874-016-0280-6>.
- Mozumder, S. I., Rutherford, M. J. ja Lambert, P. C. (2021). Estimating restricted mean survival time and expected life-years lost in the presence of competing risks within flexible parametric survival models. *BMC medical research methodology* 21, lk. 1–20. DOI: <https://doi.org/10.1186/s12874-021-01213-0>.
- Parner, E. T., Andersen, P. K. ja Overgaard, M. (2023). Regression models for censored time-to-event data using infinitesimal jack-knife pseudo-observations, with applications to left-truncation. *Lifetime Data Analysis* 29.3, lk. 654–671. DOI: <https://doi.org/10.1007/s10985-023-09597-5>.

- Prasad, R. B. ja Groop, L. (2015). Genetics of type 2 diabetes—pitfalls and possibilities. *Genes* 6.1, lk. 87–123. DOI: <https://doi.org/10.3390/genes6010087>.
- Tartu Ülikooli Eesti geenivaramu (2025). URL: <https://geenidoonor.ee/geenivaramu> (vaadatud 28.02.2025).
- Teder, K. (2023). *Teise tüübi diabeedi riski prognoosimine ja tagasiside algoritmi väljatöötamine TÜ Eesti geenivaramu andmetel*. Bakalaureusetöö. Tartu, Eesti: Tartu Ülikool. URL: <https://dspace.ut.ee/items/843a62b2-b65b-4bfd-aed3-bd437422d7e7> (vaadatud 04.02.2025).
- Therneau, T. (2024a). *Pseudo-values for survival data*. URL: <https://cran.r-project.org/web/packages/survivalVignettes/vignettes/pseudo.html> (vaadatud 16.05.2025).
- Therneau, T., Crowson, C. ja Atkinson, E. (2024). *Multi-state models and competing risks*. URL: <https://cran.r-project.org/web/packages/survival/vignettes/compete.pdf> (vaadatud 15.02.2025).
- Therneau, T. (2024b). *pseudo: Pseudo values for survival*. CRAN. URL: <https://www.rdocumentation.org/packages/survival/versions/3.8-3/topics/pseudo> (vaadatud 16.05.2025).
- World Health Organization (14. november 2024). *Diabetes*. URL: <https://www.who.int/news-room/fact-sheets/detail/diabetes> (vaadatud 26.04.2025).
- Zhang, Z. (2016). Survival analysis in the presence of competing risks. *Annals of Translational Medicine* 5.3. URL: <https://atm.amegroups.org/article/view/11637> (vaadatud 23.02.2025).

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Karmel Teder,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Absoluutrisi prognoosimine konkureerivate riskide korral: meetodite võrdlus teise tüüpi diabeedi näitel“, mille juhendaja on Krista Fischer, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Karmel Teder

21.05.2025