

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Kelly Tilga
**Pseudovaatlused elukestusanalüüsis
depressiooni ja kardiometaboolsete haiguste
vaheliste seoste hindamiseks TÜ Eesti
Geenivaramu andmete põhjal**

Kindlustus- ja finantsmatemaatika

Magistritöö (30 EAP)

Juhendaja: MSc Anastassia Kolde

TARTU 2023

**PSEUDOVAATLUSED ELUKESTUSANALÜÜSIS DEPRESSIOONI
JA KARDIOMETABOOLSETE HAIGUSTE VAHELISTE SEOSTE
HINDAMISEKS TÕ EESTI GEENIVARAMU ANDMETE PÕHJAL**

Magistritöö

Kelly Tilga

Lühikokkuvõte

Magistritöö eesmärk on uurida pseudovaatluste kasutamist elukestusanalüüsis: leides kõigepealt pseudovaatlused ning siis sobitades need mudelitesse.

Töö teoreetilises osas tehakse ülevaade klassikalisest elukestusanalüüsis, pseudovaatlustest ja nende arvutamisest.

Töö praktilises osas uuritakse depressiooni ja kardiometaboolsete haiguste vahelisi seoseid TÕ Eesti Geenivaramu andmete näitel. Võrreldakse konkureerivate riskide analüüsi klassikalise Coxi võrdeliste riskide mudeli ja pseudovaatlustega üldistatud hinnanguvõrrandite (GEE) mudelit *cloglog* linkfunktsiooniga. Uuritakse ka tõkestatud keskmise eluea seoseid kovariaatidega ja kaotatud aastaid.

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

Märksõnad: Elukestusanalüüs, pseudovaatlused, depressioon, kardiometaboolsed haigused, tõkestatud keskmine üleelamisaeg, kaotatud aastad.

**PSEUDO-OBSERVATIONS IN SURVIVAL ANALYSIS TO
EVALUATE THE RELATION BETWEEN DEPRESSION AND
CARDIOMETABOLIC DISEASES BASED ON ESTONIAN GENOME
CENTRE'S DATA**

Master thesis

Kelly Tilga

Abstract

The aim of this master thesis is to study the use of pseudo-observations in survival analysis: first finding the pseudo-observations and then fitting them in models.

The theoretical part of the thesis gives an overview of survival analysis, pseudo-observations and how to calculate pseudo-observations.

In the practical part of the thesis, the relation between depression and cardiometabolic diseases is studied using data from Estonian Genome Centre. Competing risks analysis comparison of the classical Cox proportional hazard model and generalized estimating equations with cloglog link function using pseudo-observations is made. Restricted mean survival time and years lost models and analysis is also included.

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

Key Words: Survival analysis, pseudo-observations, depression, cardiometabolic diseases, restricted mean survival time, years lost.

Sisukord

Sissejuhatus	5
1 Elukestusanalüüs	6
1.1 Olulisemad definitsioonid	6
1.2 Tsenseerimine	7
1.3 Kaplan-Meieri hinnang	8
1.4 Coxi võrdeliste riskide mudel	9
1.5 Konkureerivad riskid	10
2 Pseudovaatlused	11
2.1 Hinnang üleelamistõenäosusele	12
2.2 Tõkestatud keskmine üleelamisaeg	12
2.3 Konkureerivad riskid	14
2.4 Üldistatud hinnanguvõrrandid	15
2.5 Kaotatud aastad	17
2.6 Näited pseudovaatluste arvutamisest	19
2.6.1 Näide üleelamistõenäosusest	19
2.6.2 Näide tõkestatud keskmisest	19
2.7 Kasutatud R-i paketid	20
3 Andmete analüüs	22
3.1 Kardiometaboolsetest haigustest ja nende riskiteguritest	22
3.2 Andmete ülevaade	23
3.3 Mudelite võrdlused	25

3.4 Tõkestatud keskmine ja kaotatud aastad	31
Kokkuvõte	36
Kasutatud allikad	37
Lisa 1. Mudel 2 kaotatud aastad teist tüüpi diabeedi ja surma korral	40
Lisa 2. Mudel 2 kaotatud aastad südame isheemiatõve ja insuldi korral	41

Sissejuhatus

Käesoleva magistritöö eesmärk on uurida pseudovaatluseid elukestusanalüüsis, keskendudes nende arvutamisel ja modelleerimisel.

Magistritöö esimeses peatükis tuuakse välja peamised elukestusanalüüsi definitsioonid ja meetodid: Kaplan-Meieri hinnang, Coxi mudel ja konkureerivad riskid. Teises peatükis tutvustatakse pseudovaatlusi ja käsitletakse pseudovaatluste arvutamist. Tuuakse välja üleelamistöenäosus, tõkestatud keskmine üleelamisaeg, konkureerivad riskid, üldistatud hinnanguvõrrandid ja kaotatud aastad. Samuti tehakse kaks näidet pseudovaatluste arvutamise kohta ning tutvustatakse pseudovaatluste leidmist tarkvarapaketi R abil. Töö kolmandas osas kirjeldatakse TÜ Eesti Geenivaramu andmeid ja analüüsitakse depressiooni ja kardiometaboolsete haiguste vahelisi seoseid. Võrreldakse pseudovaatluste metoodikat klassikalise Coxi võrdeliste riskide lähenemisega konkureerivate riskide puhul ja uuritakse tõkestatud keskmist ja kaotatud aastate seost erinevate kovariaatidega.

Töö praktilises osas vaadeldakse TÜ Eesti Geenivaramu andmeid kardiometaboolsetest haigustest, millest on vaatluse all järgnevad haigused: rinnaangiin, kodade virvendus ja laperdus, südame isheemiatõbi, insult ja teist tüüpi diabeet. Uuritakse depressiooni mõju eelnimetatud kardiometaboolsetele haigustele. Töös tehakse konkureerivate riskide mudel, kus ühe konkureeriva riskina on arvestatud ka surma. Vaadeldakse nii Coxi võrdeliste riskide mudelit kui ka pseudovaatlustega üldistatud hinnanguvõrrandi (*GEE*) mudelit *cloglog* linkfunktsiooniga. Lisaks arvutatakse tõkestatud keskmine kõikide kardiometaboolsetele haigustele ja seejärel leitakse kaotatud aastad vastavalt igale kardiometaboolsele haigusele.

1 Elukestusanalüüs

Peatükk 1 ja selle alapeatükid põhinevad raamatul "*Modelling Survival Data in Medical Research*" (Collett, 2015), kui ei ole märgitud teisti.

Selles peatükis antakse ülevaade elukestusanalüüsi olulisematest mõistetest, tsenseerimisest ja erinevatest analüüsi lähenemistest.

Elukestusanalüüs on statistika haru, mis analüüsib aega kuni sündmusteni. Huvipakkuvateks sündmusteks saavad olla näiteks surm, haiguse diagnoos, kindlustusnõue, haigusest taastumisaeg.

Mõiste 'elukestusanalüüs' tuleneb biomeditsiini teadlastelt, sest kasutatavad meetodid pärinevad biomeditsiinilistest huvidest uurida suremust, patsientide üleelamisaega haiguse diagnoosist surmani. Elukestusanalüüsi esimene mäрге pärineb 17ndast sajandist ning esimese elukestusanalüüsi läbiviijaks peetakse John Graunti, kes tuletas esimese elu tabeli (inglise keeles *life table*) ning avaldas teadusartikli '*Natural and Political Observations Made Upon the Bill of Mortality*' aastal 1662. Pärast seda kui Cox avaldas oma paberi elukestusanalüüsist aastal 1972, sai alguse tänapäevane ja edasiarenenud elukestusanalüüs. (Guo, 2010)

Kuna 'aeg sündmuseni' andmetel esineb tsenseerimist, siis ei saa kasutada klassikalisi mudeleid nagu näiteks logistiline regressioon ning seeõttu eksisteeribki eraldi elukestusanalüüs. Andmete analüüsimiseks kasutatakse näiteks Kaplan-Meieri hinnangut, Coxi võrdeliste riskide mudelit või parameetrilist regressiooni.

1.1 Olulisemad definitsioonid

Selles alapeatükis tuuakse välja elukestusanalüüsi kolm olulisemat definitsiooni.

Tähistame T kui pidevat mittenegatiivset juhuslikku suurust, mis näitab indiviidi elukestust. Olgu juhusliku suuruse T kumulatiivne jaotusfunktsioon $F(t)$.

Definition. *Tõenäosus, et huvipakkuv sündmus pole toimunud enne ajahetke t , on*

antud üleelamisfunktsiooniga (survival function):

$$S(t) = P(T > t) = 1 - F(t). \quad (1)$$

Teiseks oluliseks mõisteks on riskifunktsioon.

Definition. Riskifunktsioon (hazard function) on piirväärtus kui Δt läheneb nullile ning on defineeritud järgmiselt:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (2)$$

Riskifunktsioon näitab sündmuse juhtumise tõenäosust ajahetkel, eeldades, et sündmus ei ole toimunud enne ajahetke t . (Kleinbaum ja Klein, 2005)

Definition. Kumulatiivne riskifunktsioon (cumulative hazard function) on kumulatiivne risk, et sündmus toimub ajani t ning avaldub valemiga

$$H(t) = \int_0^t h(u) du$$

1.2 Tsenseerimine

Selles alapeatükis kirjeldatakse tsenseeritud andmete eripära, esinemist ja olulisust.

Elukestusanalüüsi eripära tuleneb andmete tsenseerimisest. Tsenseerimine esineb kui andmetes on teada mingi üleelamisaeg, kuid täpne aeg on teadmata. (Kleinbaum ja Klein, 2005)

Tsenseerimine võib esineda näiteks kui sündmust ei toimu uuringu ajal, indiviid lahkub uuringust enne lõppaega. Eristatakse kolme tsenseerimise viisi: paremalt tsenseerimine, vasakult tõkestatud ja intervalltsenseerimine.

Paremalt tsenseerimine esineb, kui indiviidi üleelamisaeg on pikem kui analüüsi lõppaeg. Paremalt tsenseeritud üleelamisaeg on väiksem kui reaalne üleelamisaeg,

kuid viimane on teadmata, kuna uuring lõppes varem, kui sündmus toimus. Vasakult tsenseerimine esineb, kui reaalne üleelamisaeg on väiksem kui vaadeldud üleelamisaeg. Paremtalt tsenseerimine on tavapärasem ja esineb tihemini kui vasakult tõkestatust või intervalltsenseerimist. Reaalset üleelamisaega võib lugeda ka tsenseerituks, kui indiviid suri põhjusesse, mis ei ole uuringuga seotud. Kuid sellel on omad komplikatsioonid, sest surmapõhjus võib ikkagi kaudselt olla seotud uuritavate põhjustega, kuid pole hästi tõestatav.

Tsenseeritud vaatlused on küll puudulikud, kuid siiski olulised, sest nende eemaldamisel väheneb analüüsi olulisus.

Elukestusanalüüsis on oluline eeldus, et tsenseerimine on sõltumatu ja mitteinformatiivne (*independent and non-informative censoring*). See tähendab, et tsenseerimise aeg c ei anna informatsiooni sündmuse toimumise kohta ega millegi muu kohta, kui ainult seda, et paremtalt tsenseerimise korral $c < t$. Informatiivse tsenseerimise puhul tuleb kasutusele võtta teistsugune käsitlus, mida selle magistritöö raames ei vaadelda.

1.3 Kaplan-Meieri hinnang

Olgu andmestikus n indiviidi, kelle vaadeldavad üleelamisajad on t_1, t_2, \dots, t_n . Eeldame, et sündmuste arv on r nii, et $r \leq n$. Järjestame üleelamissajad kasvavasse järjekorda niimoodi, et $t_{(1)} < t_{(2)} < \dots < t_{(r)}$. Olgu d_j , $j = 1, 2, \dots, r$, indiviidide arv, kellel toimub sündmus ajahetkel t_j ning n_j kõikide indiviidide arv, kellel pole sündmust toimunud enne ajahetke t_j ning kaasaarvatud need, kellel varsti sündmus toimub.

Võtame nüüd eelduseks, et sündmuste esinemised on üksteisest sõltumatud. Eeldatav üleelamisfunktsioon ajahetkel t on hinnanguline üleelamistõenäosus peale ajahetke $t_{(k)}$, mis on k -s intervallis $t_{(k)}$ kuni $t_{(k+1)}$, $k = 1, 2, \dots, r$, kus $t_{(r+1)}$ on määratud ∞ . Ehk see on üleelamistõenäosus intervallis $t_{(k)}$ kuni $t_{(k+1)}$ ning järgnevates perioodides. Saame Kaplan-Meieri üleelamistõenäosuse hinnangu valemi:

$$\widehat{S}(t) = \prod_{j=1}^k \left(\frac{n_j - d_j}{n_j} \right), \quad (3)$$

kus $t_{(k)} \leq t < t_{(k+1)}$, $k = 1, 2, \dots, r$ ning $\widehat{S}(t) = 1$ kui $t < t_{(1)}$.

1.4 Coxi võrdeliste riskide mudel

Kõigepealt vaatame üldist võrdeliste riskide mudelit. Olgu olukord, kus risk sündmuse toimumiseks kindlal ajahetkel, sõltub kovariaatidest x_1, \dots, x_p .

Olgu p argumenttunnust X_1, \dots, X_p ning olgu nende väärtused võrdeliste riskide mudelis määratud vektoriga $\mathbf{x} = (x_1, \dots, x_p)'$ ja olgu $h_0(t)$ baasriskifunktsioon indiviidil, kelle muutujate väärtused on võrdsed nulliga ehk $\mathbf{x} = 0$. Riskifunktsiooni i -ndale indiviidile saab kirjutada

$$h_i(t) = \psi(\mathbf{x}_i)h_0(t), \quad (4)$$

kus $\psi(\mathbf{x}_i)$ on \mathbf{x}_i funktsioon.

Kuna $\psi(\mathbf{x}_i)$ ei saa olla negatiivne, siis võime kirjutada selle lineaar kombinatsioonina ja võtta eksponendi ehk saame valemi (4) kirjutada kujul

$$h_i(t) = \exp(\beta_1 x_{1i}, \beta_2 x_{2i}, \dots, \beta_p x_{pi})h_0(t). \quad (5)$$

Olgu andmestikus n indiviidi, millest r on kindla sündmuse toimumise ajaga ning $n - r$ paremalt tsenseeritud. Eeldame, et sündmuse toimumise ajad ei kattu. Järjestame sündmuste ajad $t_{(1)} < t_{(2)} < \dots < t_{(r)}$ nii, et j -ndas sündmuse aeg oleks $t_{(j)}$. Olgu $R(t_{(j)})$ indiviidide hulk, kes on riski all ajahetkel $t_{(j)}$, tähistame. Seega $R(t_{(j)})$ on indiviidid, kellel pole sündmust toimunud ja tsenseerimata enne aega $t_{(j)}$.

Coxi tõepärafunktsioon mudelile (5) avaldub kujul:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^r \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\boldsymbol{\beta}' \mathbf{x}_l)}. \quad (6)$$

Coxi mudeli puhul on oluline, et riskide suhted on võrdsed ning ei muutu ajas. Eelduseid saab kontrollida erinevatel viisidel nagu näiteks Schoenfeldi jääkidega, modifitseeritud Cox-Snelli jääkidega, martingaaljääkidega.

Tarkvarapaketi R saab riskide võrdelisust kontrollida funktsiooniga `cox.zph`, kus p-väärtus väiksem kui 0,05 näitab, et riskide võrdelisuse eeldus ei ole täidetud. (Therneau, 2023)

1.5 Konkureerivad riskid

Konkureerivate riskide elukestusandmete modelleerimiseks saab kasutada Coxi mudelit, parameetrilist elukestusmudelit või mudelit, mis kasutab kumulatiivset esinemissagedust (*cumulative incidence*) elukestuse asemel. Tüüpilise põhjusepõhise lähenemise idee elukestusanalüüsis konkureerivate riskide korral on teha igale sündmuse tüübile eraldi analüüs, kus ülejäänud tüüpe käsitletakse kui tsenseeritud. (Kleinbaum ja Klein, 2005)

Konkureerivate riskide puhul võtame arvesse, et sündmuseid on mitmeid. Võtame üleelamisaja T ja põhjuse C . Sellisel juhul indiviidi vaadeldav väärtus on (T, C) , kus i -nda indiviidi andmed on kujul (t_i, c_i) , $i = 1, \dots, n$, kus c_i võimalikud väärtused on $0, 1, 2, \dots, m$ ja $c_i = 0$ kui sündmus ei ole vaadeldud.

Konkureerivate riskide puhul ei ole Kaplan-Meieri hinnang sobiv, sest see ülehindab sündmuse juhtumist ning annab harva hea tõlgenduse konkureerivatele riskidele.

Pikemalt käsitletakse konkureerivaid riske alapeatükis 2.3.

2 Pseudovaatlused

Peatükk 2 ja selle alapeatükid põhinevad Anderseni ja Perme'i artiklil "*Pseudo-observations in survival analysis*" (Andersen ja Perme, 2010), kui ei ole märgitud teisiti.

Elukestusanalüüsi omapäraks on mittetäielikud ehk tsenseeritud andmed. Mittetäieliku andmestiku korral saab kasutada meetodeid, mis sisaldavad parameetrilisi või mitteparameetrilisi hinnanguid, üleelamise jaotust ning ka selle parameetrilist või poolparameetrilist regressioonianalüüsi, mis on tihti saadud riskifunktsiooni regressioonimudelite kaudu.

Täieliku andmestiku (ehk andmestiku, kus ei esine tsenseerimist) korral, saab üleelamisaega T vaadelda kõikide indiviidide korral ning standardseid meetodeid saab rakendada otse.

Tsenseeritud andmete korral on üheks võimaluseks asendada funktsioon $f(T)$ pseudoväärtusega (töös nimetatud ka kui pseudovaatlus, inglise keeles *pseudo-observation*) ning seda hakataksegi käsitlema.

Täieliku andmestiku puhul on $f(T)$ vaadeldav iga indiviidi i jaoks ning keskvärtus on leitav valemiga $E(f(T)) = \frac{1}{n} \sum_i f(T_i)$.

Vaadates nüüd aga mittetäieliku andmestikku, võtame hinnangu $\hat{\theta}$ keskvärtusele $\theta = E(f(T))$. Sellisel juhul on pseudoväärtus funktsioonile $f(T)$ i -nda indiviidi jaoks defineeritud järgmiselt:

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

kus $\hat{\theta}^{-i}$ on hinnang valimile suurusega $n-1$ (see elimineerib i -nda väärtuse andmestikus). Ehk i -ndat pseudoväärtust saab vaadelda kui i -nda indiviidi panust hinnangule $E(f(T))$ valimil suurusega n .

2.1 Hinnang üleelamistõenäosusele

Olgu paremalt tsenseeritud andmed esitatud järgnevalt, (\tilde{X}_i, D_i) , $i = 1, \dots, n$, kus \tilde{X}_i on võrdne üleelamisajaga X_i kui $D_i = 1$ ning on võrdne paremalt tsenseeritud andmetega, kui $D_i = 0$. Saame selle kokku võtta kui loendusprotsessi (*counting process*)

$$N(t) = \sum_i I(\tilde{X}_i \leq t, D_i = 1), \quad (7)$$

andes vaadeldavad sündmused lõigus $[0, t]$ ning

$$Y(t) = \sum_i I(\tilde{X}_i \geq t), \quad (8)$$

on vaadeldavate arv, kes on riski all enne ajahetke t .

Hinnang $\theta = S(t)$ on võrdne Kaplan-Meieri hinnanguga

$$\hat{S}(t) = \prod_{u \leq t} \left(1 - \frac{dN(u)}{Y(u)}\right).$$

Sellisel juhul on i -s pseudoväärtus arvutatav valemiga

$$\hat{S}_i(t) = n\hat{S}(t) - (n-1)\hat{S}^{-i}(t). \quad (9)$$

2.2 Tõkestatud keskmine üleelamisaeg

Tõkestatud keskmist saab tõlgendada kui keskmist üleelamisaega või eeldatavat elukestust kindla ajaperioodi jooksul, kus aeg on alates 0-st kuni valitud ajahetkeni. Tõkestatud keskmine üleelamisaeg on hea meetod kliinilistes uuringutes, sest see annab tähendusliku viisi tõlgendada üleelamise erinevust gruppide vahel. Tõkestatud keskmise eeliseks on see, et sellel ei ole osasid piiranguid, näiteks ei pea arvestama riskide proportsionaalsusega nagu peab Coxi võrdeliste riskide mudeli korral. Tõkestatud keskmine ja keskmine üleelamisaeg on sarnased, aga tõkestatud

keskmise eeliseks on see, et see on usaldusväärsem kui andmetes esineb tsenseerimist, sest keskmise üleelamisaeg eeldab, et kõik sündmused on teada ja tsenseerimist ei esine, sest muidu võib tulla mitteusaldusväärne tulemus. (Han ja Jung, 2022)

Olgu T juhuslik suurus, kas arv, vektor või protsess ja olgu parameeter θ antud valemiga $\theta = E(f(T))$.

Olgu alguspunktiks üleelamisaeg T ja τ tõkestatud keskmine

$$\mu_\tau = E(T \wedge \tau)$$

ühe fikseeritud väärtuse jaoks $\tau > 0$. Seega $f(T) = f_\tau(T) = T \wedge \tau$ ja $\theta = \mu_\tau$. Kuna

$$T \wedge \tau = \int_0^{T \wedge \tau} 1 dt = \int_0^\tau I(T > t) dt,$$

saame, et $E(T \wedge \tau) = \int_0^\tau S(t) dt$.

Olgu integraali all Kaplan-Meieri hinnang

$$\hat{\mu}_\tau = \int_0^\tau \hat{S}^{-i}(t) dt. \quad (10)$$

Saame, et valem (10) hindab τ -tõkestatud keskmist $\theta = E(T \wedge \tau)$.

Sellisel juhul i -s pseudoväärtus on arvatav valemiga

$$\hat{\mu}_{\tau i} = n \int_0^\tau \hat{S}(t) dt - (n-1) \int_0^\tau \hat{S}^{-i}(t) dt = \int_0^\tau \hat{S}_i(t) dt, \quad (11)$$

kus $\hat{S}_i(t)$ on antud valemiga (9).

2.3 Konkureerivad riskid

Kõigepealt vaatame konkureerivate riskide mudelit, millel on kaks sündmuse võimalust. See on mitme olekuga mudel, mis võtab väärtuseid 0, 1 või 2 ehk 0 kui indiviid on tsenseeritud, 1 kui indiviidil toimub sündmus põhjuse number 1 tõttu enne ajahetke t ja 2 kui indiviidil toimub sündmus põhjuse number 2 tõttu enne ajahetke t . Olgu sündmuse põhjuse indikaator $D = 1, 2$ ja elukestus T . Kumulatiivne esinemissagedus põhjusele 1 ajahetkel t_0 on antud valemiga:

$$C_1(t_0) = E(I(T(t_0) = 1)).$$

Kuid ajahetki võib olla mitmeid (t_1, t_2, \dots, t_k) või vaadates kogu aega $t > 0$, saame kumulatiivse esinemissageduse valemiks põhjusele 1, $D = 1, 2$

$$C_1(t) = E(I(T \leq t, D = 1)) = \int_0^t S(u-) \alpha_1(u) du,$$

kus $S(u) = P(X < u)$ on üldine üleelamistõenäosus $P(X(u) = 0)$ ning $\alpha_1(u)$ on põhjusepõhine risk sündmusele põhjusest 1. Seega saame

$$f(X(\cdot)) = f_t(T(\cdot)) = I(T(t) = 1)$$

ja $\theta = C_1(t)$.

Vaatame nüüd elukestusi (\tilde{X}_i, D_i) , $i = 1, 2, \dots, n$, kus \tilde{X}_i tähistab paremalt tsenseeritud sündmuste aegu ning $D_i = 0$ näitab tsenseerimist ning $D_i = j$, kui $j = 1, 2$, näitab, kas indiviidil i toimus sündmus põhjusel j . Saame ajavahemikus $[0, t]$ loendada põhjuse j , $j = 1, 2$, sündmuseid valemiga

$$N_j(t) = \sum_i I(\tilde{X}_i \leq t, D_i = j).$$

Saame vaadelda indiviidide arvu, kellel on risk sündmust kogeda ajahetkel t vale-

miga

$$Y(t) = \sum_i I(\tilde{X}_i \geq t).$$

Kumulatiivne esinemissagedus põhjusele 1 on hinnatav Nelson-Aaleni hinnanguga

$$\hat{C}_1(t) = \int_0^t \hat{S}(u-) d\hat{A}_1(u),$$

kus $\hat{A}_1(t) = \int_0^t \frac{1}{\bar{Y}(u)} dN_1(u)$ on Nelson-Aaleni hinnang kumulatiivsele põhjusepõhisele riskile $A_1(t) = \int_0^t \alpha_1(u) du$ põhjuse 1 sündmustele. Saamegi leida konkureerivate riskide mudeli puhul põhjusele 1 i -nda pseudoväärtuse valemiga

$$\hat{C}_{1i}(t) = \int_0^t \hat{S}_i(u-) d\hat{A}_1(u).$$

Või siis üldistatult põhjusele j valemiga

$$\hat{C}_{ji}(t) = \int_0^t \hat{S}_i(u-) d\hat{A}_j(u).$$

2.4 Üldistatud hinnanguvõrrandid

Pseudovaatlused pakuvad elukestusanalüüsi lähenemistele mitmesuguseid mudeleid, asendades mittetäieliku vaatluse väljundi ning sobitades selle kasutades näiteks üldistatud hinnanguvõrrandeid (*GEE - generalized estimating equations*).

Olgu X skalaar, vektor $X = (X_1, \dots, X_k)$, juhuslik suurus või stohhastiline protsess $X(t)$. Olgu θ antud valemiga $\theta = E(f(X))$ ja olgu siin f , ja seega ka θ , mitmemõõtmeline ning viimaks olgu $Z = (Z_1, \dots, Z_p)$ kovariaatide vektor fikseeritud ajaga. Nüüd vaatame mudelit $E(f(X|Z))$ jaoks. Vaadates üldistatud lineaarset mudelit koos antud linkfunktsiooniga g saame

$$g(E(f(X)|Z)) = \beta_0 + \sum \beta_j Z_j.$$

Näide selle mudeli kasutamisest on näiteks Coxi võrdeliste riskide regressiooni mudel, kus f on ühemõõtmeline ja vaadeldakse ajahetke t_0 , kus $f(X) = I(X > t_0)$ ja seega $\theta = S(t_0)$. Võttes nüüd linkfunktsiooniks *cloglog* funktsiooni $g(x) = \text{cloglog}(x) = \log(-\log(1 - x))$ ja saame

$$\log(-\log(S(t_0))) = \log A_0(t_0) + \sum_j \beta_j Z_j, \quad (12)$$

kus $A_0(t) = \int_0^t \alpha_0(u) du$ on kumulatiivne baasrisk Coxi mudelis $\alpha(t|Z) = \alpha_0(t) \exp(\sum \beta_j Z_j)$.

Saame valemite (12) laiendada ka mitmetele ajahetkedele t_1, \dots, t_k ning saame valemi

$$\log(-\log(S(t_l))) = \log A_0(t_l) + \sum_j \beta_j Z_j,$$

kus $l = 1, \dots, k$ ja $f(X) = f_{t_1}(X), \dots, f_{t_k}(X)$ on k -mõõtmeline.

Või siis saab valemite (12) rakendada kõikide ajahetkede $t > 0$ jaoks

$$\log(-\log(S(t))) = \log A_0(t) + \sum_j \beta_j Z_j.$$

Tsenseeritud andmestiku puhul saab kasutada pseudovaatlusi, et mudeleid sobitada. Indiviidi i jaoks on pseudovaatlus $\hat{\theta}_i$ k -mõõtmeline ehk $\hat{\theta}_i = (\hat{\theta}_{ij}, j = 1, \dots, k)$. See tähendab, et näiteks Coxi mudelis kõikidel ajahetkedel on väärtus kõikide vaadeldud sündmuste jaoks. Siis $\hat{\theta}_{ij}$ on $f_{t_j}(X_i)$ ja mudel on $g(E(f_{t_j}(X_i)|Z_i)) = \beta^T Z_{ij}^*$. Siin vektor Z_{ij}^* sisaldab kõiki ajahetki ja kovariaate Z_i ehk $Z_{ij}^* = (I(t_l = t_j), l = 1, \dots, k; Z_i)$.

Parameetrite β hinnangud $k + p$ -mõõtmises esinevad järgmise valemiga

$$\sum_i \left(\frac{\partial}{\partial \beta} g^{-1}(\beta^T Z_i^*) \right)^T V_i^{-1} (\hat{\theta}_i - g^{-1}(\beta^T Z_i^*)) = \sum_i U_i(\beta) = U(\beta) = 0. \quad (13)$$

Siin V_i on *working covariance matrix*, mida tihti kasutatakse *id*-na (Andersen, 2013)

ja $g^{-1}(\beta^T Z_i^*)$ on lühem versioon k -vektorist elementidega $g^{-1}(\beta^T Z_{ij}^*)$, $j = 1, \dots, k$.

Olgu

$$I(\beta) = \sum_i \left(\frac{\partial g^{-1}(\beta^T Z_i^*)}{\partial \beta} \right)^T V_i^{-1} \left(\frac{\partial g^{-1}(\beta^T Z_i^*)}{\partial \beta} \right),$$

ning

$$\widehat{\text{var}}(U(\beta)) = \sum_i U_i(\hat{\beta})^T U_i(\hat{\beta}),$$

siis saame $\hat{\beta}$ dispersiooni hinnata *sandwich* hinnanguga, mis avaldub järgmisel kujul

$$\widehat{\text{var}}(\beta) = I(\hat{\beta})^{-1} \widehat{\text{var}}(U(\hat{\beta})) (I(\hat{\beta})^{-1}).$$

2.5 Kaotatud aastad

See alapeatükk põhineb artiklil *Estimating restricted mean survival time and expected life-years lost in the presence of competing risks within flexible parametric survival models* (Mozumder, Rutherford ja Lambert, 2021).

Vaatame kõigepealt tõkestatud keskmist üleelamisega konkureerivatele riskidele.

Tõkestatud keskmine üleelamisega näitab keskmist üleelamisega või keskmist elatud aega indiviidil alates ajast 0 kuni defineeritud hetkeni τ . Tõkestatud keskmine kovariaatide puudumisel on valemiga

$$\mu(\tau) = E(\min(T, \tau)) = \int_0^\tau S(u) du, \quad (14)$$

kus $S(t)$ on üleelamisfunktsioon.

Saame leida eeldatavad kaotatud aastad enne aega τ järgmise valemiga

$$L(0, \tau) = \tau - E(\min(T, \tau)) = \tau - \int_0^\tau S(u) du. \quad (15)$$

Konkureerivate riskide puhul saab kaotatud aastate koguarvu, $L(0, \tau)$, leida läbi

kaotatud aastate iga sündmuse k korral. Olgu

$$S(t) = 1 - \sum_{k=1}^K F_k(t),$$

siis tõkestatud keskmist üleelamisaega valemis (14) saab leida iga põhjusepõhise kumulatiivse esinemissageduse funktsiooni järgmise valemiga

$$\mu(\tau) = E(\min(T, \tau)) = \int_0^\tau S(u) du = \int_0^\tau 1 - \sum_{k=1}^K F_k(u) du = \tau - \sum_{k=1}^K \int_0^\tau F_k(u) du.$$

Valemi (15) saame esitada integraalina iga põhjusepõhise kumulatiivse esinemissageduse funktsioonile järgmiselt

$$L(0, \tau) = \tau - \int_0^\tau S(u) du = \sum_{k=1}^K \int_0^\tau F_k(u) du,$$

mida saab tõlgendada ka kui tõkestatud keskmine sündmuse toimumise aeg (*restricted mean failure time, RMFT*). Ja siit saame eeldatava kaotatud aastate arvu leidmiseks järgmise valemi

$$L_k(0, \tau) = \int_0^\tau F_k(u) du,$$

mis annab eeldatavad kaotatud aastad põhjusele k enne aega τ .

Kasutades lähenemist valemiga (13) saame leida eeldatavad kaotatud aastad enne aega τ põhjusest k valemiga

$$\hat{\theta}_k^i = nL_k(0, \tau) - (n-1)L_k^{-i}(0, \tau),$$

kui $i = 1, \dots, n$. Seda arvutab ka R-i pakett *pseudo* funktsiooniga *pseudoyl*. (Andersen, 2013)

2.6 Näited pseudovaatluste arvutamisest

Selles alapeatükis tuuakse näiteid pseudovaatluste kohta vaadeldes üleelamistõenäosust ja tõkestatud keskmist.

2.6.1 Näide üleelamistõenäosusest

Võtame suvalised vaatlused ja suvalised sündmused. Olgu antud 10 vaatlust, millest kaks on tsenseeritud. Arvutame pseudoväärtused kõigil juhtudel olenemata kas sündmus toimus või ei toimunud. Tabelis on välja toodud pseudoväärtused elukesetusele pikkusega 47 ja 240. Tabelist on näha, et $t = 47$ korral algavad pseudoväärtused ühest ja peale seda vähenevad kuni nullini. Kuid $t = 240$ korral algavad väärtused ühest ning seejärel tõusevad natuke üle ühe ning jõudes 240ni langevad alla nulli.

Tabel 1: Näide üleelamistõenäosusest

Vaatlus	Sündmus	Pseudovaatlused	
		$t = 47$	$t = 240$
15	1	1	1
33	1	1	1
47	0	1	1
53	1	0,857	1,024
139	1	0,714	1,048
196	1	0,571	1,071
240	1	0,429	-0,071
244	0	0,429	-0,071
403	1	0,214	-0.036
507	1	0	0

2.6.2 Näide tõkestatud keskmisest

Võtame täpselt samad andmed, mis olid näites 1 ning kasutame tõkestatud keskmist. Valime juhuslikult kolm tõket 70, 200 ja 550. Näeme, et $t = 70$ korral hakkavad pseudoväärtused peale valitud t saabumist muutuma samadeks. Sama mustrit on

näha $t = 200$ korral, kus pseudoväärtused enne 200 kõiguvad ning peale 200 jäävad samale tasemele. Võttes aga t piisavalt suure või seda mitte defineerides, näeme, et sellist mustrit ei teki. Samuti näeme, et olenevalt tsenseerimisest sõltub pseudo-väärtus: kui sündmus toimus, siis pseudoväärtus on lähedal vaatlusega, kui vaatlus oli tsenseeritud, siis pseudovaatlus ei pruugi olla lähedane number vaatlusega.

Tabel 2: Näide tõkestatud keskmisest

Vaatlus	Sündmus	Pseudovaatlused		
		t = 70	t = 200	t = 550
15	1	15	15	15
33	1	33	33	33
47	0	67,57	169,71	284,71
53	1	50,57	33,55	14,38
139	1	70,4	133,88	114,71
196	1	70,4	200,38	181,21
240	1	70,4	205,05	232,55
244	0	70,4	205,05	483,38
403	1	70,4	205,05	362,05
507	1	70,4	205,05	604,71

2.7 Kasutatud R-i paketid

Antud magistritöös kasutatakse andmete analüüsiks ja modelleerimiseks tarkvara-paketti R. Pseudovaatluseid saab leida paketi *pseudo*, mis koosneb neljast funktsioonist: *pseudosurv*, *pseudomean*, *pseudoci*, *pseudoyl*. (Maja Pohar Perme ja Gerster, 2017)

Selles töös kasutatakse funktsioone *pseudoci*, *pseudomean* ja *pseudoyl*.

Funktsioon *pseudomean(time, event, tmax)* on pseudoväärtuste arvutamiseks tõkestatud keskmise meetodil. Argumendiks *event* on binaarne tunnus sündmuse kohta ning *tmax* on tõkestatud keskmise ajapunkt. Funktsioon tagastab pseudoväärtuste vektori tõkestatud keskmise jaoks. (Maja Pohar Perme ja Gerster, 2017)

Funktsioon $pseudoci(time, event, tmax)$ on konkureerivate riskide arvutamiseks kumulatiivsel esinemissagedusel. Võrreldes $pseudomean$ -iga, on $pseudoci$ korral $event$ konkureeriva riski staatuse määramiseks: 0 tähendab tsenseerimist ja $1, \dots, k$ tähistavad sündmuseid. Argument $tmax$ on ajahetkede vektor, millel pseudovaatlused arvutatakse. Argumentides $time$ ja $event$ ei tohi eksisteerida puuduvaid väärtusi. (Maja Pohar Perme ja Gerster, 2017)

Funktsioon $pseudoyl$ on kaotatud aastate leidmiseks. Sellel on argumendid $time$ ja $event$ sarnaselt $pseudoci$ -ga, kus sündmus tähistab konkureeriva riski staatust. Funktsioon tagastab vektorite listi vastavalt põhjustele. (Maja Pohar Perme ja Gerster, 2017)

Pseudovaatluste mudeldamiseks kasutatakse paketi $geepack$ funktsiooni $geese$. See funktsioon sobitab üldistatud hinnanguvõrrandi mudeli. Argument $mean.link$ on keskmiste linkfunktsiooni defineerimiseks, määrates selleks $identity$ linkfunktsiooni, saab leida riskide erinevust, võttes log saab leida riskide suhet ($risk\ ratio$) ning $cloglog$ korral riskimäärade suhet ($hazard\ ratio$). Kui $geese$ mudelisse panna pseudo-väärtused ja kasutada $cloglog$ linkfunktsiooni, siis see peaks lähenema Coxi mudelite hinnangutega. (Halekoh, Højsgaard ja Yan, 2006)

Samuti eksisteerivad R-is paketid $prodlm$ ja $eventglm$, millega saab ka arvutada ja modelleerida pseudovaatlusi ning ka pakett $fastpseudo$, mis sisaldab ainult üht funktsiooni, millega saab leida tõkestatud keskmist, kuid selle magistritöö raames neid pakette prooviti, kuid ei kasutatud.

3 Andmete analüüs

Selles magistritöös kasutakse TÜ Eesti Geenivaramu andmeid, et analüüsida depressiooni ja kardiometaboolsete haiguste vahelisi seoseid pseudovaatluste abil. Töös võrreldakse klassikalise Coxi võrdeliste riskide mudeli tulemusi ja pseudovaatluste põhjal tehtud mudeleid. Samuti tehakse tõkestatud keskmise üleelamisaja ja kaotatud aastate analüüs. Andmete modelleerimiseks ja arvutuste tegemiseks kasutatakse tarkvarapaketti R. Antud töös kasutatakse samasid andmeid, mis Laura Birgit Luitva magistritöös aastal 2022 (Luitva, 2022).

3.1 Kardiometaboolsetest haigustest ja nende riskiteguritest

Rinnaangiini teine nimetus on stenokardia, mille põhitunnuseks on valu või pigistustunne rinnus. See võib esile tulla näiteks suurel füüsilisel pingutusel või tugeva emotsionaalse pinge all. Rinnaangiin on südame isheemiatõve leebem vorm. (Südamekeskus, 2004)

Kodade virvendus ja laperdus on südame rütmihäired, mille põhitunnuseks on kiire ja ebaregulaarne südame rütm. (Kaljusaar ja K.Sukles, 2019) Kodade virvenduse ja laperduse põhiliseks riskiteguriks on vanem vanus, kuid samuti ka ebatervislik elustiil. (Järving *et al.*, 2021)

Südame isheemiatõbi on haigus, kui südamelihaste normaalne hapniku ja toitainetega varustamine on häiritud ning mille tagajärjel langeb südame normaalne töövoime. (Teeäär *et al.*, 2019)

Insult ehk peajuinfarkt on ajuveresoonte ummistus või lõhkemine, mille tagajärjel aju ei saa piisavalt hapnikku ega toitaineid. Nagu ka teiste kardiometaboolsete haiguste korral, on ka insuldi korral peamisteks riskifaktoriteks ebatervislik eluviis ehk näiteks suitsetamine ja liigne alkoholi tarbimine. Samuti on riskifaktoriteks ka kõrge vererõhk, südame haigused ja diabeet. (ITK, 2020)

Teist tüüpi diabeet on energiaainevahetushäire, mis tuleneb insuliini vähesest tootmisest või resistentsusest. Teist tüüpi diabeedi puhul on oluliseks riskifaktoriks pärilikkus, kuid lisaks on veel ülekaalulisus, kõrge vererõhk ja veresuhkur. (ITK, 2021)

Uuringutes on tulnud välja, et depressiooniga inimestel on suurem tõenäosus saada kardiometaboolne haigus. Vastupidine suhe on ka välja toodud, et kardiometaboolse haigusega inimene võib kergemini depressiooni sattuda. Depressiooniga indiviididel võivad tekkida ebatervislikumad eluviisid ning seetõttu saada kardiometaboolse haiguse. (Hare *et al.*, 2013)

3.2 Andmete ülevaade

Töös vaadeldakse järgmiseid kardiometaboolseid haiguseid: rinnaangiin, kodade virvendus ja laperdus, südame isheemiatõbi, insult, teist tüüpi diabeet. Andmestik on kokku 92 611 indiviidi, kes ei ole suguluses ja liitumishetkel olid vähemalt 18-aastased. Andmestikus esineb veel indiviidide sugu, kehamassiindeks (normaalkaalus $KMI \leq 25$, ülekaalus $KMI > 25$), haridustase (kuni põhiharidus, keskharidus, kõrgharidus), suitsetamise staatus (mitte kunagi, endine, praegune), alkoholi tarbimise staatus (mitte kunagi, endine, praegune) ning füüsiline aktiivsus (tegeleb tervisespordiga, ei tegele tervisespordiga).

Töös uuritakse baastaseme depressiooni mõju kardiometaboolsete haiguste riskile. Vaatlusperiood on alates 01.01.2011 kuni 31.12.2021 ehk 11 aastat. Baastaseme depressioon selle töö lõikes tähendab, et depressioon on diagnoositud vahemikus 2006 kuni 2010. Andmestikust on eemaldatud järgmised indiviidid: kellel diagnoositi depressioon enne 2006. aastat, kes surid enne 2011. aastat, kellel esimest korda diagnoositi kõik viis kardiometaboolset haigust enne 2011. aastat, kellel on mingi kardiometaboolne haigus diagnoositud enne 2011. aastat, kellel on kõik kovariaatide andmed olemas ja teada enne 2011. aastat.

Ajaskaala kasutatakse vanust. Erinevates uuringutes on välja toodud, et kohati

on mõistlikum kasutada vanust ajaskaalana, kui näiteks teha elukestusanalüüs vanemaast uuringugrupist (Lamarca *et al.*, 1998). Veel on välja toodud, et kasutades vanust ajaskaalana võib anda isegi täpsemad tulemused, kuna see lubab vaadata samas riskirühmas olevaid indiviide. Tihti ei pruugi tulemustes väga suuri erinevusi olla võrreldes kasutades uuringus olemise aega ajaskaalana või siis vanust ajaskaalana. (Canchola *et al.*, 2003)

Järgmisena teeme ülevaate kovariaatidega andmestikule suurusega 18 243 indiviidi. Selles andmestikus on naisi kokku 12 196 ja mehi 6047. Keskmine vanus aastal 2011 on 41,04. Tabelist 3 näeme, et kõige rohkem esineb rinnaangiini, kovariaatidega andmestikus 802 ja selle järel teist tüüpi diabeet, 684 juhtu. Kõige vähem esineb insult, 184 juhtu. Terveid inimesi on kovariaatidega andmestikus 15 052. Andmetike suuruste erinevused tulenevad sellest, et kovariaatidega andmestiku jaoks sai eemaldatud kõik vaatlused, mis olid puudulikud ning kovariaatide andmed ei olnud teada enne 2011. aastat.

Tabel 3: Haiguste arvud

	Kogu andmestik	Kovariaatidega andmestik
Rinnaangiin	2811	802
Kodade virvendus ja laperdus	1763	521
Südame isheemiatõbi	1610	462
Insult	500	184
Teist tüüpi diabeet	2591	684
Surm	923	538

Tabelist 4 näeme, et depressioonita indiviide on rohkem kui depressiooniga. Depressioonita indiviidide seas on rohkem naisi kui mehi, keskharidusega indiviidide arv on suurem kui kuni põhiharidusega kui ka kõrgharidusega indiviidide arv, mitte kunagi suitsetamise staatusega on kõige rohkem indiviide võrreldes endise ja praeguse suitsetamise staatusega, praeguse alkoholi tarbijaid on palju rohkem kui mitte kunagi alkoholi tarbinud ja endiseid tarbijaid ning füüsiliselt aktiivseid indiviide on veidi rohkem kui mitteaktiivseid. Depressiooniga indiviidide seas on jagunemine

sarnane.

Tabel 4: Riskifaktorite jagunemine

	Depressioonita n = 16 232	Depressiooniga n = 2011
Sugu		
Mees	5612	435
Naine	10 620	1576
Haridustase		
Kuni põhiharidus	2155	277
Keskharidus	9687	1190
Kõrgharidus	4390	544
Suitsetamise staatus		
Mitte kunagi	9159	1046
Endine	1792	259
Praegune	5281	706
Alkoholi tarbimine		
Mitte kunagi	1504	194
Endine	479	84
Praegune	14 249	1733
Füüsiline aktiivsus		
Ei ole aktiivne	7019	905
On aktiivne	9213	1106

3.3 Mudelite võrdlused

Selles alapeatükis vaatame Coxi võrdeliste riskide mudeli ja GEE mudeli *cloglog* linkfunktsiooniga võrdlusi.

Mudelid on tehtud konkureerivate riskidega, kus riskideks on rinnaangiin, kodade virvendus ja laperdus, südame isheemiatõbi, insult, teist tüüpi diabeet ja samuti on konkureeriva riskina arvesse võetud surm. Pseudovaatluste korral ei ole kummaski mudelis arvestatud vasakult tõkestatust, kuna pseudovaatluste korral ei ole vasakult tõkestamine oluline. Teeme võrdluse Coxi võrdeliste riskide mudeli arvestades vasakult tõkestamist ja seda mittearvestades ning pseudovaatlustega mudeli vahel. Selles töös kasutatakse olulisuse nivood $\alpha = 0,05$. Pseudovaatlused arvutati aja-

hetkedel 20-75 viieaastaste sammudena. Seejärel pseudovaatluste mudeldamiseks kasutatakse üldistatud hinnanguvõrrandi ehk GEE mudelit *cloglog* linkfunktsiooniga.

Konkureerivate riskide korral tehakse kaks mudelit:

- Mudel 1: mudelis on depressioon ja sugu (baastasemeks on invidiidid, kes on depressioonita ja mehed);
- Mudel 2: mudelis on depressioon ja peamised riskifaktorid: haridustase, suitsetamise staatus, alkoholi tarbimise staatus, füüsiline aktiivsus, sugu ja kehamassiindeks (baastasemeks on invidiidid, kes on depressioonita, keskhari- dusega, mitte kunagi suitsetanud, mitte kunagi alkoholi tarbinud, füüsiliselt mitteaktiivsed, mehed ja normaalkaalus).

Selles töös on nii mudeli 1 kui ka mudeli 2 korral kasutatud kovariaatidega and- mestikku. Põhjuseks on see, et paketi *pseudo* funktsioonid ei töötanud suurema andmestiku korral või puudus selleks piisavalt võimekas arvuti.

Tabelis 5 on välja toodud võrdlus Coxi võrdeliste riskide mudel ja pseudovaatluste- ga GEE *cloglog* mudel. Vasakpoolne Coxi mudeli tulp on arvatud arvesse võttes vasakult tõkestamist ja paremalt tsenseerimist ning on märgitud lõpuga *VP*, kuid parempoolses Coxi mudeli tulbas on arvesse võetud vaid paremalt tsenseeritust ning märgitud *P*.

Näeme tabelist 5, et nii Coxi mudelite korral kui pseudovaatlustega GEE mudeli *cloglog* linkfunktsiooniga hinnangud on sarnased. Näeme, et võrreldes Coxi võrde- liste riskide VP mudelit ja pseudovaatlustega mudelit, siis rinnaangiini korral on depressiooni hinnang veidi vähenenud, olles Coxi VP mudeli korral 1,55 ja pseu- dovaatluste korral 1,48. See tähendab, et depressiooniga invidiidil on pseudovaat- lustega mudeli põhjal 1,48 korda suurem risk saada rinnaangiin kui depressioonita invidiidil ning naistel on 1,25 ($\approx \frac{1}{0,8}$) korda väiksem risk saada rinnaangiin kui meestel. Kodade virvenduse ja laperduse korral on Coxi VP mudelis depressiooni

hinnang 1,12, kuid pseudovaatlustega mudelis on kerge ülehinnang ning hinnanguks on 1,17, kuid mõlema mudeli korral ei ole depressioon statistiliselt oluline. Depressioon ei ole statistiliselt oluline veel südame isheemiatõve, insuldi ja teist tüüpi diabeedi korral. Statistilise olulisuse erinevus soo korral tuleb teist tüüpi diabeedi puhul, sest Coxi mudeli puhul ei ole soo korral riskide proportsionaalsuse eeldus täidetud. Proportsionaalsuse eeldus soo korral täidetud ainult insuldi ja surma korral ning südame isheemiatõve korral kui vasakult tõkestatust ei olnud arvesse võetud. Surma korral on Coxi VP mudelis depressiooni hinnang 1,46 ning see on statistiliselt oluline ning pseudovaatluste korral on hinnang 1,6.

Tabel 5: Mudel 1 Coxi konkureerivate riskide mudeli ja pseudovaatlustega mudeli võrdlus

	Coxi mudel VP		Coxi mudel P		PO	
	$\exp(\hat{\beta}_i)$	p-väärtus	$\exp(\hat{\beta}_i)$	p-väärtus	$\exp(\hat{\beta}_i)$	p-väärtus
Rinnaangiin						
Depressioon	1,55	$4,3 \cdot 10^{-5}$	1,56	$5,9 \cdot 10^{-6}$	1,48	$1,4 \cdot 10^{-4}$
Sugu: naine	0,68	$6,7 \cdot 10^{-7}$	0,69	$1,3 \cdot 10^{-6}$	0,8	$3,1 \cdot 10^{-3}$
Kodade virvendus						
Depressioon	1,12	0,42	1,21	0,16	1,17	0,31
Sugu: naine	0,56	$2,4 \cdot 10^{-10}$	0,58	$1,2 \cdot 10^{-9}$	0,55	$2,7 \cdot 10^{-9}$
Südame isheemiatõbi						
Depressioon	1,1	0,51	1,18	0,25	1,22	0,23
Sugu: naine	0,47	$1,6 \cdot 10^{-14}$	0,52	$5,3 \cdot 10^{-12}$	0,55	$9,5 \cdot 10^{-9}$
Insult						
Depressioon	0,77	0,34	0,83	0,49	0,62	0,17
Sugu: naine	0,63	$3,5 \cdot 10^{-3}$	0,61	$1,4 \cdot 10^{-3}$	0,64	$7,7 \cdot 10^{-3}$
Teist tüüpi diabeet						
Depressioon	1,02	0,83	1,15	0,24	1,09	0,44
Sugu: naine	0,78	$2,6 \cdot 10^{-3}$	0,82	0,02	0,88	0,1
Surm						
Depressioon	1,46	$3,8 \cdot 10^{-3}$	1,5	$8,1 \cdot 10^{-4}$	1,6	$1,8 \cdot 10^{-3}$
Sugu: naine	0,34	0	0,35	0	0,39	0

Coxi konkureerivate riskide mudeli tulemused ja pseudovaatluste mudelite tulemused mudel 2 korral on välja toodud tabelites 6, 7, 8 ja 9. Hinnangud on mudeli-

tes sarnased, kuid erinevused Coxi mudelite ja pseudovaatluste mudelitel võivad tuleneda Coxi võrdeliste riskide mudeli puhul riskide proportsionaalsuse eelduse mittetäitmisest.

Näeme tabelist 6, et rinnaangiini korral on nii Coxi mudelites kui ka pseudovaatluste mudelis depressioon statistiliselt oluline, Coxi VP mudelis on depressiooni hinnang 1,56, kuid pseudovaatluste mudelis 1,46. Mõlemas mudelis on statistiliselt olulised veel sugu ja kehamassiindeks. Näeme, et kui haridustase on kuni põhiharidus, siis Coxi VP mudeli korral ei ole see tunnus oluline, kuid Coxi P mudeli puhul on see statistiliselt oluline hinnanguga 0,61 ning pseudovaatluste korral on see samuti statistiliselt oluline ning hinnanguga 0,63. Praegune suitsetamise staatus ei ole statistiliselt oluline Coxi VP mudelis ja pseudovaatluste mudelis, kuid on oluline Coxi P mudelis.

Tabel 6: Mudel 2 võrdlused rinnaangiini korral

	Coxi mudel VP		Coxi mudel P		PO	
	$\exp(\hat{\beta}_i)$	p-väärtus	$\exp(\hat{\beta}_i)$	p-väärtus	$\exp(\hat{\beta}_i)$	p-väärtus
Rinnaangiin						
Depressioon	1,56	$1,3 \cdot 10^{-5}$	1,56	$5,8 \cdot 10^{-6}$	1,46	$3,4 \cdot 10^{-4}$
Haridus: kuni põhiharidus	0,92	0,5	0,61	$4,6 \cdot 10^{-5}$	0,63	$2,8 \cdot 10^{-4}$
Haridus: kõrgharidus	0,91	0,25	0,87	0,09	0,91	0,31
Suitsetamine: endine	0,84	0,15	0,94	0,59	0,91	0,51
Suitsetamine: praegune	0,94	0,49	1,27	$6,2 \cdot 10^{-3}$	1,1	0,23
Alkohol: endine	1,2	0,37	1,14	0,51	1,5	0,07
Alkohol: praegune	0,85	0,2	0,97	0,8	1,07	0,62
Füüsiline aktiivsus	0,95	0,53	1,09	0,23	1,16	0,09
Sugu: naine	0,7	$8,0 \cdot 10^{-7}$	0,7	$8,9 \cdot 10^{-6}$	0,8	$2,8 \cdot 10^{-3}$
Kehamassiindeks: üle	1,4	$1,6 \cdot 10^{-5}$	1,23	$7,1 \cdot 10^{-3}$	1,2	$9,6 \cdot 10^{-3}$

Tabelist 7 näeme veel, et kodade virvenduse ja laperduse korral ei ole depressioon statistiliselt oluline mitte ühegi mudeli puhul. Coxi VP mudeli ja pseudovaatluste mudeli korral on statistiliselt olulised järgmised tunnused: suitsetamine: endine, sugu: naine, kehamassiindeks: üle. See tähendab, et näiteks ülekaalulistel on Coxi VP mudeli põhjal 1,54 korda suurem risk saada kodade virvendus ja laperdus kui

normaalkaalus olevatel indiviididel ning pseudovaatluste mudeli järgi 1,2 korda. Naistel on 1,7 ($\approx \frac{1}{0,59}$) korda väiksem risk saada kodade virvendus ja laperdus kui meestel.

Tabel 7: Mudel 2 võrdlused kodade virvenduse korral

	Coxi mudel VP		Coxi mudel P		PO	
	$\exp(\hat{\beta}_i)$	p-väärtus	$\exp(\hat{\beta}_i)$	p-väärtus	$\exp(\hat{\beta}_i)$	p-väärtus
Kodade virvendus						
Depressioon	1,13	0,39	1,22	0,16	1,13	0,42
Haridus: kuni põhiharidus	1,12	0,36	0,79	0,06	0,76	0,09
Haridus: kõrgharidus	0,75	0,01	0,69	$8,5 \cdot 10^{-4}$	0,78	0,07
Suitsetamine: endine	1,29	0,048	1,36	0,017	1,52	0,012
Suitsetamine: praegune	1,25	0,06	1,64	$1,2 \cdot 10^{-5}$	1,31	0,02
Alkohol: endine	0,78	0,37	0,7	0,2	0,47	0,23
Alkohol: praegune	0,87	0,37	1,02	0,86	1,09	0,69
Füüsiline aktiivsus	0,96	0,72	1,09	0,34	1,98	0,88
Sugu: naine	0,59	$1,9 \cdot 10^{-7}$	0,64	$4,4 \cdot 10^{-6}$	0,62	$6,4 \cdot 10^{-6}$
Kehamassiindeks: üle	1,54	$2,6 \cdot 10^{-5}$	1,42	$3,1 \cdot 10^{-4}$	1,2	0,03

Tabelist 8 näeme, et südame isheemiatõve korral ei ole depressioon statistiliselt oluline mitte üheski mudelis. Olulisteks tunnusteks kõikides mudelites on suitsetamine: praegune ja sugu: naine. Coxi VP mudeli korral on praegustel suitsetajatel 1,54 korda suurem risk saada südame isheemia kui mitte kunagi suitsetajatel ning pseudovaatluste mudelite korral on see näitaja 1,69. Naistel on võrreldes meestega 2 ($=\frac{1}{0,5}$) korda väiksem risk ning pseudovaatluste mudeli põhjal ligikaudu 1,8.

Tabelist 8 on veel näha, et insuldi korral ei ole mitte ühegi mudeli korral depressioon statistiliselt oluline. Ainsana on naissugu oluline kõikides mudelites hinnaguga 0,65, nii Coxi VP mudelis kui ka pseudovaatluste mudelis, ehk naistel on 1,5 ($\approx \frac{1}{0,65}$) korda väiksem risk kui meestel.

Tabelist 9 näeme, et teist tüüpi diabeedi puhul ei ole depressioon statistiliselt oluline. Näiteks on kehamassiindeks oluline teist tüüpi diabeedi korral, ülekaalulistel on Coxi VP mudeli põhjal 4,7 korda suurem risk saada teist tüüpi diabeet ja pseudovaatluste põhjal 2,7.

Tabel 8: Mudel 2 võrdlused südame isheemiatõve ja insuldi korral

	Coxi mudel VP		Coxi mudel P		PO	
	$\exp(\hat{\beta}_i)$	p-väärtus	$\exp(\hat{\beta}_i)$	p-väärtus	$\exp(\hat{\beta}_i)$	p-väärtus
Südame isheemiatõbi						
Depressioon	1,09	0,55	1,18	0,25	1,25	0,22
Haridus: kuni põhiharidus	1,25	0,11	0,85	0,24	0,69	0,03
Haridus: kõrgharidus	1,05	0,65	1,0	0,99	0,96	0,8
Suitsetamine: endine	1,04	0,78	1,01	0,93	1,01	0,94
Suitsetamine: praegune	1,54	$2,1 \cdot 10^{-4}$	1,92	$1,3 \cdot 10^{-8}$	1,69	$1,7 \cdot 10^{-5}$
Alkohol: endine	1,18	0,47	1,18	0,48	0,75	0,2
Alkohol: praegune	0,67	$7,4 \cdot 10^{-3}$	0,79	0,11	0,87	0,14
Füüsiline aktiivsus	0,91	0,35	1,02	0,84	0,94	0,26
Sugu: naine	0,5	$8,4 \cdot 10^{-11}$	0,57	$3,9 \cdot 10^{-8}$	0,55	$2,5 \cdot 10^{-7}$
Kehamassiindeks: üle	1,17	0,11	1,13	0,2	1,07	0,5
Insult						
Depressioon	0,77	0,36	0,83	0,51	0,54	0,07
Haridus: kuni põhiharidus	1,36	0,15	0,95	0,79	1,13	0,58
Haridus: kõrgharidus	0,88	0,51	0,88	0,5	0,95	0,82
Suitsetamine: endine	0,93	0,78	1,04	0,87	1,24	0,5
Suitsetamine: praegune	1,31	0,15	1,63	$8,1 \cdot 10^{-3}$	1,51	0,015
Alkohol: endine	1,35	0,42	1,38	0,36	0,73	0,6
Alkohol: praegune	0,75	0,24	0,88	0,57	0,87	0,63
Füüsiline aktiivsus	0,73	0,06	0,81	0,2	0,78	0,16
Sugu: naine	0,65	0,012	0,66	0,012	0,65	$9,2 \cdot 10^{-3}$
Kehamassiindeks: üle	1,05	0,78	0,97	0,81	0,8	0,25

Surma korral on statistiliselt olulised tunnused kõikides mudelites järgmised: depressioon, haridus: kuni põhiharidus, suitsetamine: praegune, sugu: naine ja kehamassiindeks: üle. Depressiooniga indiviididel on Coxi VP mudeli järgi 1,45 korda suurem risk surra kui depressioonita ning pseudovaatluste mudeli korral 1,4. Praegustel suitsetajatel on ligi 3 korda suurem risk surra kui mitte kunagi suitsetanud indiviididel. Kuni põhiharidusega indiviididel on Coxi VP mudeli põhjal 1,84 korda suurem risk surra kui keskharitutel ning pseudovaatluste korral 1,48. Naistel on Coxi VP mudeli põhjal ligi 2,3 ($\approx \frac{1}{0,44}$) korda väiksem risk surra kui meestel ning pseudovaatluste korral 1,9 korda väiksem.

Tabel 9: Mudel 2 võrdlused teist tüüpi diabeedi ja surma korral

	Coxi mudel VP		Coxi mudel P		PO	
	$\exp(\hat{\beta}_i)$	p-väärtus	$\exp(\hat{\beta}_i)$	p-väärtus	$\exp(\hat{\beta}_i)$	p-väärtus
Teist tüüpi diabeet						
Depressioon	1,04	0,74	1,19	0,15	1,12	0,42
Haridus: kuni põhiharidus	1,34	0,014	0,84	0,13	0,8	0,14
Haridus: kõrgharidus	0,63	$1,1 \cdot 10^{-5}$	0,62	$2,1 \cdot 10^{-6}$	0,58	$3,2 \cdot 10^{-5}$
Suitsetamine: endine	1,11	0,38	1,14	0,29	1,19	0,26
Suitsetamine: praegune	1,23	0,03	1,71	$3,5 \cdot 10^{-9}$	1,4	$8,7 \cdot 10^{-4}$
Alkohol: endine	1,03	0,89	1,03	0,89	1,08	0,8
Alkohol: praegune	0,87	0,27	1,02	0,86	1,06	0,74
Füüsiline aktiivsus	0,76	$1,1 \cdot 10^{-3}$	0,92	0,28	0,82	0,03
Sugu: naine	0,89	0,18	0,91	0,3	0,98	0,83
Kehamassiindeks: üle	4,7	0	3,9	0	2,7	0
Surm						
Depressioon	1,45	$4,5 \cdot 10^{-3}$	1,57	$6,5 \cdot 10^{-4}$	1,4	0,04
Haridus: kuni põhiharidus	1,84	$1,4 \cdot 10^{-8}$	1,37	$2,8 \cdot 10^{-3}$	1,48	$1,6 \cdot 10^{-3}$
Haridus: kõrgharidus	0,82	0,09	0,77	0,03	0,79	0,18
Suitsetamine: endine	1,32	0,06	1,33	0,053	1,32	0,24
Suitsetamine: praegune	2,79	0	3,67	0	2,9	0
Alkohol: endine	1,2	0,42	1,08	0,74	2,24	0,02
Alkohol: praegune	0,85	0,27	0,98	0,91	1,3	0,23
Füüsiline aktiivsus	0,82	0,04	0,96	0,64	0,83	0,1
Sugu: naine	0,44	0	0,49	$4 \cdot 10^{-14}$	0,52	$7 \cdot 10^{-10}$
Kehamassiindeks: üle	0,79	$8,5 \cdot 10^{-3}$	0,76	$1,8 \cdot 10^{-3}$	0,65	$6,3 \cdot 10^{-4}$

3.4 Tõkestatud keskmine ja kaotatud aastad

Tabelites 10, 11, 12, tabelis 13 lisas 1 ja tabelis 14 lisas 2 on p-väärtused märgitud tärniga (*) kui väärtus on väiksem kui 0,05.

Kõigepealt leiti tõkestatud keskmine olenemata kardiometaaboolsest haigusest (k.a surm) ja siis kaotatud aastad vastavalt kardiometaaboolsele haigusele, et näha kardiometaaboolse haiguse seost tõkestatud keskmise elukestusega. Vaadeldavateks ajapunktideks sai valitud $t = 30, 50, 70$.

Tabelist 10 näeme, et $t = 30$ korral võib depressiooniga indiviid elada kardiometaaboolsete haiguste vaba elu keskmiselt 0,0081 aastat isegi kauem. Depressiooniga

indiviidid võivad kaotada keskmiselt 0,002 haigustevaba aastat vähem kui depressioonita indiviid rinnaangiini tõttu ning teist tüüpi diabeedi tõttu 0,0026 aastat vähem.

Tabelist 10 näeme, et $t = 50$ korral kaotatakse depressioonile 0,079 kardiometaboolsete haiguste vaba aastat. Depressiooniga indiviidid võivad kaotada keskmiselt 0,07 haigustevaba aastat rohkem kui depressioonita indiviidid rinnaangiini tõttu, kodade virvenduse ja laperduse tõttu 0,00025 aastat vähem, südame isheemiatõve korral 0,0065 haigustevaba aastat rohkem, insuldi tõttu 0,02 aastat vähem, teist tüüpi diabeedi tõttu 0,0032 aastat rohkem ning surma tõttu 0,022 aastat rohkem.

Tabelist 10 näeme, et $t = 70$ korral kaotatakse depressioonile 1,18 kardiometaboolsete haiguste vaba aastat. Depressiooniga indiviidid võivad kaotada keskmiselt 0,59 haigustevaba aastat rohkem kui depressioonita indiviidid rinnaangiini tõttu, kodade virvenduse ja laperduse tõttu 0,11 aastat rohkem, südame isheemiatõve korral 0,072 haigustevaba aastat rohkem, insuldi tõttu 0,0078 aastat vähem, teist tüüpi diabeedi tõttu 0,18 aastat rohkem ning surma tõttu 0,31 aastat rohkem.

Näeme, et naised võivad elada keskmiselt 2,3 kardiometaboolsete haiguste vaba aastat rohkem kui mehed. Naised võivad kaotada keskmiselt 0,47 haigustevaba aastat vähem kui mehed rinnaangiini tõttu, 0,41 aastat vähem kodade virvenduse ja laperduse tõttu, südame isheemiatõve tõttu 0,39 aastat vähem, insuldi tõttu 0,12 aastat vähem, teist tüüpi diabeedi tõttu 0,27 aastat vähem, ning surma tõttu 0,64 aastat vähem.

Eelnevalt seletatud tulemustest näeme, et näiteks $t = 30$ korral on tulemused väga väikesed ning kohati ebaloogilised. Loogilisemad tulemused tulevad $t = 70$ korral ning tunnustel ja haigustel on suurem seos. See tuleneb sellest, et kardiometaboolsed haigused on pigem vanemate inimeste haigused ning ebatervislikud eluviisid on elukäiku rohkem mõjutanud.

Tabelist 11 näeme, et mudel 2 korral olenemata haigusest võib $t = 30$ korral depressiooniga indiviid elada keskmiselt 0,009 haigustevaba aastat isegi rohkem, kuid

Tabel 10: Mudel 1 tõkestatud keskmine ja kaotatud aastad

	Tõkestatud keskmine		
	t=30	t= 50	t=70
Kõik haigused			
Depressioon	0,0081 *	-0,079	-1,18 *
Sugu: naine	0,011 *	0,39 *	2,3 *
	Kaotatud aastad		
	t=30	t= 50	t=70
Rinnaangiin			
Depressioon	-0,002 *	0,07	0,59 *
Sugu: naine	-0,0037 *	-0,11 *	-0,47 *
Kodade virvendus ja laperdus			
Depressioon	-0,0008	-0,0025	0,11
Sugu: naine	-0,00095	-0,0064 *	-0,41 *
Südame isheemiatõbi			
Depressioon	-0,00024	0,0065	0,072
Sugu: naine	-0,00056	-0,031 *	-0,39 *
Insult			
Depressioon	-0,00066	-0,02 *	-0,078
Sugu: naine	0,00036	-0,0097	-0,12 *
Teist tüüpi diabeet			
Depressioon	-0,0026 *	0,0032	0,18
Sugu: naine	-0,0016	-0,061 *	-0,27 *
Surm			
Depressioon	-0,0021 *	0,022	0,31 *
Sugu: naine	-0,0047 *	-0,2 *	-0,64 *

$t = 50$ korral 0,07 aastat vähem ning $t = 70$ korral isegi 1,07 haigustevaba aastat vähem. Tabelitest 11, 12, tabelist 13 lisas 1 ja tabelist 14 lisas 2 näeme, et depressiooniga invidiidid võivad $t = 70$ kaotada keskmiselt 0,59 haigustevaba aastat rohkem kui depressioonita invidiidid rinnaangiini tõttu, kodade virvenduse ja laperduse tõttu 0,1 aastat rohkem, südame isheemiatõve korral 0,06 haigustevaba aastat rohkem, insuldi tõttu 0,08 aastat vähem, teist tüüpi diabeedi tõttu 0,15 aastat rohkem ning surma tõttu 0,26 aastat rohkem.

Tabelist 11 näeme veel, et $t = 70$ korral võib praegune suitsetaja elada kardio-metaboolsete haiguste vaba elu 1,97 aastat vähem kui mitte kunagi suitsetanud

indiviidid. Naised võivad elada keskmiselt 1,82 haigustevaba aastat rohkem kui mehed ning ülekaalulised võivad elada keskmiselt 0,9 haigustevaba aastat vähem kui normaalkaalus olevad indiviidid.

Tabel 11: Mudel 2 tõkestatud keskmine ja kaotatud aastad rinnaangiini korral

	Tõkestatud keskmine					
	t=30		t= 50		t=70	
Kõik haigused						
Depressioon	0,009	*	-0,07		-1,07	*
Haridus: kuni põhiharidus	-0,016	*	-0,16	*	0,85	*
Haridus: kõrgharidus	0,006	*	0,14	*	0,96	*
Suitsetamine: endine	0,006	*	-0,04		-0,32	
Suitsetamine: praegune	-0,007	*	-0,22	*	-1,97	*
Alkohol: endine	-0,0026		0,19		-0,19	
Alkohol: praegune	-0,004		0,047		0,12	
Füüsiline aktiivsus	-0,007	*	-0,07	*	-0,1	
Sugu: naine	0,008	*	0,31	*	1,82	*
Kehamassiindeks: üle	0,003		-0,06		-0,9	*
	Kaotatud aastad					
	t=30		t= 50		t=70	
		p		p		p
Rinnaangiin						
Depressioon	-0,0018	*	0,067		0,59	*
Haridus: kuni põhiharidus	0,0024		0,0456		-0,42	*
Haridus: kõrgharidus	-0,00048		0,0015		-0,09	
Suitsetamine: endine	-0,00004		0,018		-0,014	
Suitsetamine: praegune	0,003	*	0,047	*	0,26	*
Alkohol: endine	0,003		-0,041		0,31	
Alkohol: praegune	0,0009	*	-0,024		0,08	
Füüsiline aktiivsus	0,0013		0,024		0,14	
Sugu: naine	-0,0028	*	-0,1	*	-0,42	*
Kehamassiindeks: üle	-0,00005		-0,004		0,17	

Tabelist 11 näeme veel, et naised võivad kaotada keskmiselt 0,42 haigustevaba aastat vähem kui mehed rinnaangiini tõttu. Praegused suitsetajad võivad kaotada keskmiselt 0,26 haigustevaba aastat rohkem kui mitte kunagi suitsetanud indiviidid rinnaangiini tõttu.

Tabelist 12 näeme, et $t = 70$ korral võivad praegused suitsetajad kaotada keskmiselt 0,19 haigustevaba aastat rohkem kui mitte kunagi suitsetanud indiviidid

kodade virvenduse ja laperduse tõttu ning naised võivad kaotada keskmiselt 0,36 haigustevaba aastat vähem kui mehed. Kuni põhiharidusega invidiidid võivad kaotada keskmiselt 0,26 ning kõrgharidusega invidiidid keskmiselt 0,19 haigustevaba aastat vähem kui keskharidusega invidiidid kodade virvenduse ja laperduse tõttu.

Tabel 12: Mudel 2 kaotatud aastad kodade virvenduse ja laperduse korral

	Kaotatud aastad					
	t=30	p	t= 50	p	t=70	p
Kodade virvendus ja laperdus						
Depressioon	-0,0008	*	-0,002		0,1	
Haridus: kuni põhiharidus	0,0002		-0,027		-0,26	*
Haridus: kõrgharidus	-0,0012		-0,012		-0,19	*
Suitsetamine: endine	-0,0003		-0,021		0,22	
Suitsetamine: praegune	0,0008		0,01		0,19	*
Alkohol: endine	-0,0009		-0,025		-0,31	
Alkohol: praegune	-0,00018		-0,014		-0,035	
Füüsiline aktiivsus	0,0016		0,017		0,1	
Sugu: naine	-0,00052		-0,063	*	-0,36	*
Kehamassiindeks: üle	0,0011		0,1		0,07	

Lisas 1 olevast tabelist 13 ja lisas 2 olevast tabelist 14 saab mudel 2 korral näha kaotatud aastaid teist tüüpi diabeedi, surma, insuldi ja südame isheemiatõve korral.

Seega tulemustest saime, et $t = 30$ korral on hinnangud väga väikesed ning kohati võivad olla ebaloogilised, kuid $t = 50$ ja $t = 70$ korral tulemused loogilisemad ning hinnangud on informatiivsemad, sest kardiometaboolsed haigused esinevad pigem vanemate inimeste seas ning ebatervislikud eluviisid nagu näiteks suitsetamine ja ebatervislik toitumine on olulised riskifaktorid.

Kokkuvõte

Töö eesmärk oli uurida pseudovaatluste kasutamist elukestusanalüüsis ning TÜ Eesti Geenivaramu andmete põhjal uurida depressiooni ja kardiometaboolsete haiguste, nagu rinnaangiin, kodade virvendus ja laperdus, südame isheemiatõbi, insult ja teist tüüpi diabeet, seost. Töös kasutati konkureerivate riskide mudeleid, mille hulgas, lisaks eelnevalt mainitud kardiometaboolsetele haigustele, oli konkureeriva riskina arvestatud ka surma. Tehti Coxi võrdeliste riskide mudelid ning pseudovaatlustega üldistatud hinnanguvõrrandi mudelid *cloglog* linkfunktsiooniga. Saadud tulemusi võrreldi ning analüüsiti. Viidi läbi ka tõkestatud keskmise eluea ja kaotatud aastate analüüs.

Tulemuste analüüsist selgus, et Coxi võrdeliste riskide mudeli ja pseudovaatlustega GEE mudeli *cloglog* linkfunktsiooniga riskitegurite hinnangud on sarnased. Saadi, et depressioon on oluline riskitegur rinnaangiini ja surma korral. Depressiooni korral on ligikaudu 1,5 korda suurem risk saada rinnaangiin ning depressiooni tõttu on ligikaudu 1,6 korda suurem risk surra.

Tõkestatud keskmise eluea ja kaotatud aastate analüüsist selgust, et mudelis, kus on vaadeldud ainult depressiooni ja sugu, võivad depressiooniga iniviidid $t = 70$ korral elada ligi 1,2 kardiometaboolsete haiguste vaba aastat vähem kui depressioonita iniviidid. Depressioonile võivad iniviidid keskmiselt kaotada 0,59 haigustevaba aastat rohkem kui depressioonita iniviidid rinnaangiini tõttu ning 0,31 haigustevaba aastat rohkem surma tõttu.

Tõkestatud keskmise analüüsist selgus, et oluline on ajapunktide valik ning seos uuritava teemaga. Kardiometaboolsed haigused esinevad pigem vanematel inimestel ja seega ajapunkti valik $t = 30$ andis väga väikesed hinnangud, kuid $t = 70$ korral olid hinnangud informatiivsemad.

Seega saab analüüsi tulemuste põhjal öelda, et depressiooni saab arvestada osade kardiometaboolsete haiguste riskifaktoriks.

Kasutatud allikad

- Andersen, P. K. (detsember 2013). “Decomposition of number of life years lost according to causes of death”. *Statistics in Medicine* 32. DOI: [10.1002/sim.5903](https://doi.org/10.1002/sim.5903).
- Andersen, P. K. ja M. P. Perme (2010). “Pseudo-observations in survival analysis”. *Statistical Methods in Medical Research* 19, lk. 71–99. URL: https://journals.sagepub.com/doi/pdf/10.1177/0962280209105020?casa_token=i581ASH7R3IAAAAA:tjT2t8GemLbgMA83He7cnd6QqJ7Dch1K7j_axgx_aFH9YPzFifkTAKVkvXfWGFwaW5tJIBu8tW-BbA.
- Canchola, A. J., S. L. Stewart, L. Bernstein, D. W. West, R. K. Ross, D. Deapen, R. Pinder, P. Reynolds, W. Wright, H. Anton-Culver, D. Peel, A. Ziogas, ja P. L. Horn-Ross (märts 2003). *Cox Regression Using Different Time-Scales*. URL: https://www.lexjansen.com/wuss/2003/DataAnalysis/i-cox_time_scales.pdf.
- Collett, D. (2015). *Modelling Survival Data in Medical Research*. 3. väljaanne. Bristol: CRC Press.
- Guo, S. (2010). *Survival Analysis. Pocket Guides to Social Work Research Methods*. New York: Oxford University Press.
- Halekoh, Ulrich, Søren Højsgaard ja Jun Yan (2006). “The R Package gee-pack for Generalized Estimating Equations”. *Journal of Statistical Software* 15/2, lk. 1–11.
- Han, K. ja I. Jung (mai 2022). “Restricted Mean Survival Time for Survival Analysis: A Quick Guide for Clinical Researchers”. *Korean Journal of Radiology* 5, lk. 495–499. URL: <https://doi.org/10.3348/kjr.2022.0061>.
- Hare, D. L., S. R. Toukhsati, P. Johansson ja T. Jaarsma (november 2013). “Depression and cardiovascular disease: a clinical review”. *European He-*

- art Journal* 35.2, lk. 1365–1372. URL: <https://doi.org/10.1093/eurheartj/eh462>.
- ITK (jaanuar 2020). *Insult*. Vaadatud: 12.05.2023. URL: <https://www.itk.ee/patsiendile/patsiendi-infomaterjalid/haigused/insult>.
- (jaanuar 2021). *Diabeet*. Vaadatud: 12.05.2023. URL: <https://www.itk.ee/patsiendile/patsiendi-infomaterjalid/haigused/diabeet>.
- Järving, Kristin Ly, Tuuli Teeäär, Märt Elmet, Eve Laane ja Urmet Arus (mai 2021). *Kodade virvendusarütmia*. Vaadatud: 12.05.2023. URL: <https://www.kliinikum.ee/patsiendiinfo-andmebaas/kodade-virvendusarutmia/>.
- Kaljusaar, H. ja K.Sukles (november 2019). *Kodade virvendusarütmia*. Vaadatud: 12.05.2023. URL: <http://tromboos.ee/patsiendile/haigused-ja-seisundid/kodade-virvendusarutmia/>.
- Kleinbaum, D. G. ja M. Klein (2005). *Survival Analysis. A Self-Learning Text*. 2. väljaanne. New York: Springer.
- Lamarca, R., J. Alonso, G. Gómez ja A. Muñoz (1998). “Left-truncated data with age as time scale: an alternative for survival analysis in the elderly population”. *Journal Gerontology* 53A, lk. M337–M343. URL: [doi:10.1093/gerona/53a.5.m337](https://doi.org/10.1093/gerona/53a.5.m337).
- Luitva, Laura Birgit (2022). “Mitme olekuga mudelid depressiooni ja kardio-metaboolsete haiguste vaheliste seoste hindamiseks TÜ Eesti Geenivaramu andmete põhjal”. Magistritöö. Tartu Ülikool.
- Maja Pohar Perme ja Mette Gerster (2017). *pseudo: Computes Pseudo-Observations for Modeling*. R package version 1.4.3. URL: <https://CRAN.R-project.org/package=pseudo>.
- Mozumder, S. I., M.J. Rutherford ja P.C. Lambert (märts 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 52, lk. 201–213. URL: <https://doi.org/10.1186/s12874-021-01213-0>.

Südamekeskus, Regionaalhaigla (2004). *Stenokardia*. Vaadatud: 12.05.2023.

URL: <https://www.sydamekeskus.ee/sÃijdamehaigused/stenokardia>.

Teeäär, Tuuli, Ülle Miitel, Elviira Vool, Annika Elmet, Märt Elmet ja Marit

Aasaru (veebbruar 2019). *Krooniline südamepuudulikkus*. Vaadatud: 12.05.2023.

URL: <https://www.kliinikum.ee/patsiendiinfo-andmebaas/krooniline-sudamepuudulikkus/>.

Therneau, Terry M (2023). *A Package for Survival Analysis in R*. R package

version 3.5-5. URL: <https://CRAN.R-project.org/package=survival>.

Lisa 1. Mudel 2 kaotatud aastad teist tüüpi diabeedi ja surma korral

Tabel 13: Mudel 2 kaotatud aastad teist tüüpi diabeedi ja surma korral

	Kaotatud aastad					
	t=30	p	t= 50	p	t=70	p
Teist tüüpi diabeet						
Depressioon	-0,003	*	0,0009		0,15	
Haridus: kuni põhiharidus	0,0014		0,06		-0,17	
Haridus: kõrgharidus	-0,002	*	-0,07	*	-0,5	*
Suitsetamine: endine	-0,002	*	-0,003		0,08	
Suitsetamine: praegune	0,0009		0,03		0,38	*
Alkohol: endine	0,0004		-0,06		-0,07	
Alkohol: praegune	0,003	*	0,03		0,05	
Füüsiline aktiivsus	0,0004		-0,009		-0,05	
Sugu: naine	-0,001		-0,04		-0,09	
Kehamassiindeks: üle	0,00003		0,098	*	0,97	*
Surm						
Depressioon	-0,002	*	0,015		0,26	*
Haridus: kuni põhiharidus	0,008		0,085	*	0,25	*
Haridus: kõrgharidus	-0,0014		-0,035	*	-0,15	*
Suitsetamine: endine	-0,0016		-0,01		0,04	
Suitsetamine: praegune	0,003		0,1	*	0,76	*
Alkohol: endine	0,0003		-0,02		0,11	
Alkohol: praegune	-0,0006		-0,02		-0,02	
Füüsiline aktiivsus	0,003		0,008		-0,05	
Sugu: naine	-0,004		-0,08	*	-0,48	*
Kehamassiindeks: üle	-0,003	*	-0,05	*	-0,26	*

Lisa 2. Mudel 2 kaotatud aastad südame isheemiatõve ja insuldi korral

Tabel 14: Mudel 2 Kaotatud aastad südame isheemiatõve ja insuldi korral

	Kaotatud aastad					
	t=30	p	t= 50	p	t=70	p
Südame isheemiatõbi						
Depressioon	-0,0003		0,005		0,06	
Haridus: kuni põhiharidus	0,001		0,0023		-0,24	*
Haridus: kõrgharidus	0,0002		-0,015		-0,039	
Suitsetamine: endine	-0,0002		-0,016		0,046	
Suitsetamine: praegune	0,0004		0,025	*	0,28	*
Alkohol: endine	-0,00009		0,0007		0,21	
Alkohol: praegune	0,0002		0,02		-0,12	
Füüsiline aktiivsus	0,00006		0,006		-0,02	
Sugu: naine	-0,0004		-0,023	*	0,36	*
Kehamassiindeks: üle	-0,00002		0,018		0,04	
Insult						
Depressioon	-0,0006		-0,02	*	-0,08	
Haridus: kuni põhiharidus	0,003		-0,007		-0,02	
Haridus: kõrgharidus	-0,0002		-0,011		-0,004	
Suitsetamine: endine	-0,0009		-0,005		0,04	
Suitsetamine: praegune	-0,0009		0,08		0,09	*
Alkohol: endine	0,0002		-0,03		-0,07	
Alkohol: praegune	0,0008		-0,03		-0,08	
Füüsiline aktiivsus	0,001		0,0009		-0,03	
Sugu: naine	0,0004		-0,01		-0,11	*
Kehamassiindeks: üle	-0,0008		-0,024	*	-0,08	*

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Kelly Tilga

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose Pseudovaatlused elukestusanalüüsis depressiooni ja kardiometaboolsete haiguste vaheliste seoste hindamiseks TÜ Eesti Geenivaramu andmete põhjal, mille juhendaja on Anastassia Kolde, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Kelly Tilga

16.05.2023