

VÄHKI HAIGESTUMISE RISKI HINDAMINE VANUSE, SOO JA AASTA JÄRGI

Viktoria Kirpu
analüütik

Näited eelnevatest uuringutest ja analüüsidesest



- ◆ Keiu Paapsi (TAI) magistr töö "Laste vähi teatamise kvaliteet Eesti Vähiregistris"
- ◆ Ivar-Endrik Eiche magistr töö „Eesnäärmevähi haigestumus- ja suremustrendid Eestis“.
- ◆ Tervise Arengu Instituudi teadlased: regulaarsed uuringud, ülevaated ja artiklid vähielulemuse ja -haigestumuse kohta Eestis

Eesmärgid



Käesoleva uuringu eripära: Kasutada riskimudelite loomisel kõiki olemasolevaid andmeid (aastatest 1968-2017).



Urida:

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

Andmete kirjeldus



Vähiregister

Pahaloomulised kasvaja
(RHK10 kood "C00–C97").

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

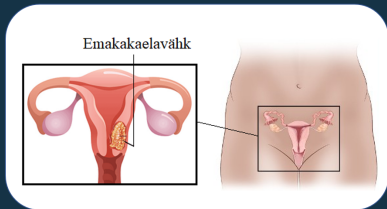
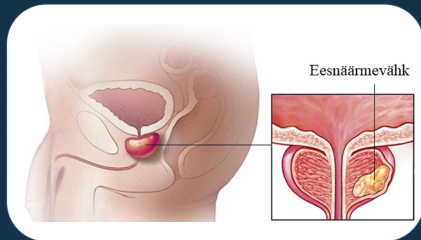
Statistikaamet

1. RV02111: Rahvastik soo ja vanuse järgi, 1. jaanuar (1950–1999)
2. RV0212: Rahvastik aasta alguses ja aastakeskmise rahvaarv soo ja vanuse järgi
 - ◆ inimeste sugu („Mees“, „Naine“)
 - ◆ inimeste vanus („0“, „1“, . . . , „84“, „85+“)
 - ◆ aasta (1968–2017)

Analüüsitud vähipaikmed



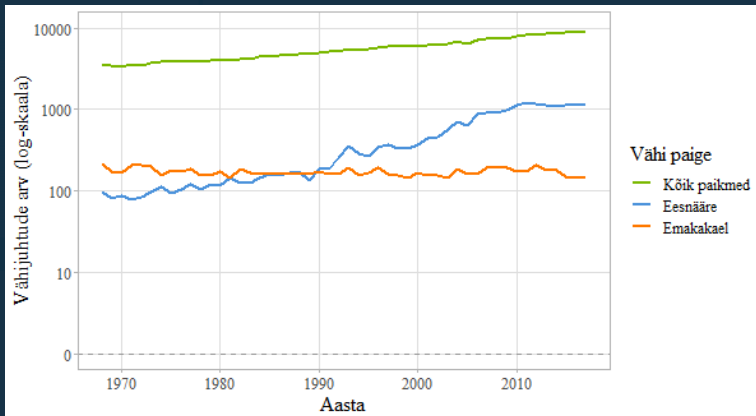
Kõik paikmed kokku



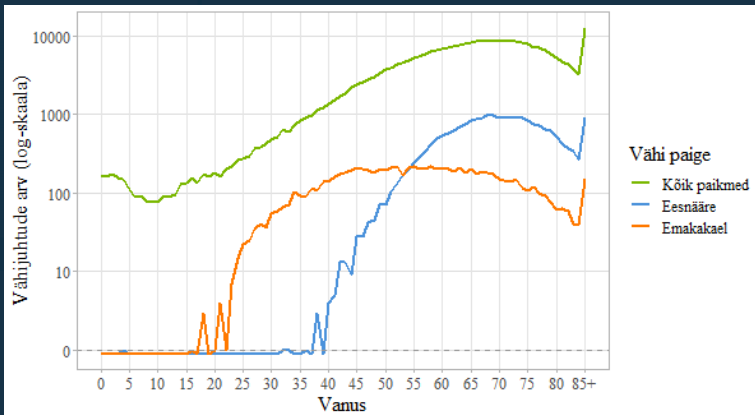
Miks?

1. Aktuaalsus
2. Soov luua unikaalseid mudeleid
3. Erinevad käitumistrendid

Vähijuhtude arv aastas



Vähijuhtude arv vanuse järgi



Mudeli valik



Mudel loodi kasutades 1968.-2016. aasta andmeid.

Kasutatud on Poissoni mudelit ja ülehajuvuse korral kvaasi-Poissoni mudelit.



$$E(y_i) = e^{\mathbf{x}'_i \boldsymbol{\beta}} = e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}$$



Grupeeritud andmete korral on vähki haigestumise riski hinnang:

$$\frac{E(y_i)}{n_i} = e^{\mathbf{x}'_i \boldsymbol{\beta}}.$$

Vektorile $\boldsymbol{\beta}$ on leitud suurima tõepära hinnang.

Mudeli hindamine



◆ Tõepärafunktsioon: $L(\beta) = \prod_{i=1}^n \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}$



◆ Log-tõepärafunktsioon:



$$L(\beta) = \ln \left(\prod_{i=1}^n \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \right) \stackrel{\mu_i = n_i e^{x_i' \beta}}{=} \sum_{i=1}^n y_i x_i' \beta + \sum_{i=1}^n n_i e^{x_i' \beta} + Const$$

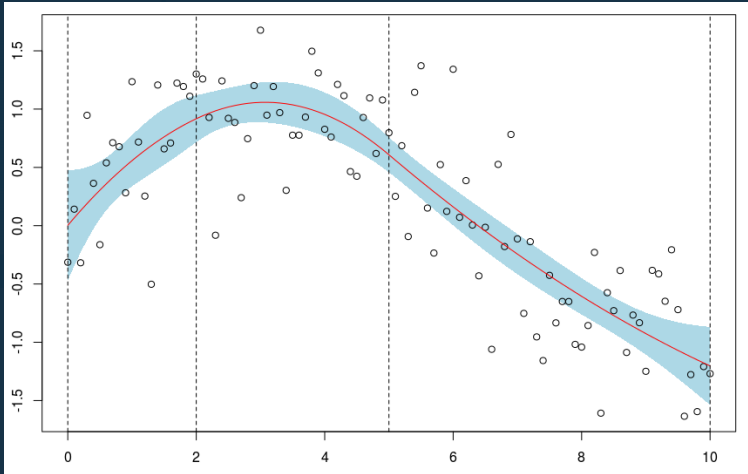


◆ Skoorifunktsioon: $s(\beta) = \frac{\partial l(\beta)}{\partial \beta} \stackrel{\mu_i = n_i e^{x_i' \beta}}{=} \sum_{i=1}^n (y_i - \mu_i) x_i$

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

$$s(\beta) = 0 \implies \sum_{i=1}^n (y_i - n_i e^{x_i' \beta}) x_i = 0 \quad \widehat{D(\hat{\beta})} = \left(\sum_{i=1}^n \hat{\mu}_i x_i x_i' \right)^{-1}$$

Splainidest



Splainid



Olgu antud muutuja x murdepunktid ξ_1, \dots, ξ_K . Üldine mudeli kuju:

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

Antud magistritöös on tunnused:

x := Vanus, kus murdepunktid on ξ_1, \dots, ξ_n

y := Aasta, kus murdepunktid on η_1, \dots, η_m

z := Sugu

Murdepunktide arvu määramiseks on kasutatud Akaike informatsioonikriteeriumit (AIC).



Seega mudel omandab kuju:

$$s(x, y, z) = \beta_0 + \beta_1 z + \beta_2 x + \beta_3 x^2 + \beta_4 x^3$$

$$+ \sum_{i=1}^n \beta_{4+i} (x - \xi_i)^3 \cdot I(x - \xi_i \geq 0) + \beta_{5+n} y$$

$$+ \beta_{6+n} y^2 + \beta_{7+n} y^3 + \sum_{i=1}^m \beta_{7+n+i} (y - \eta_i)^3 \cdot I(y - \eta_i \geq 0)$$

$$+ \beta_{8+n+m} (x \times z) + \beta_{9+n+m} (x^2 \times z) + \beta_{10+n+m} (x^3 \times z)$$

$$+ \sum_{i=1}^n \beta_{10+n+m+i} ((x - \xi_i)^3 \times z) \cdot I(x - \xi_i \geq 0)$$

$$+ \beta_{11+2n+m} (y \times z) + \beta_{12+2n+m} (y^2 \times z) + \beta_{13+2n+m} (y^3 \times z)$$

$$+ \sum_{i=1}^m \beta_{13+2n+m+i} ((y - \eta_i)^3 \times z) \cdot I(y - \eta_i \geq 0)$$

x := Vanus

y := Aasta

z := Sugu

Grupeeritud
andmete tõttu:

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

Andmete usaldusintervallid


 Hinnatud on riski ehk tõenäosust haigestuda.

Seega haigestunute arv $y \sim B(n, \theta)$.

 Mudelite töö uurimiseks on kasutatud Clopper–Pearson'i täpseid usaldusintervallid:


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

Kasvamis-, kahanemis- ja käänukohtade leidmine

 Kogu töö teeb ära R funktsioon `Estimable`.
Kontrolliti hüpoteese:

 H_0 : Tegemist on käänukohaga

 H_1 : Tegemist on kasvu- või kahanemiskohaga

 Poissoni jaotuse korral on kasutatud Waldi testi ja Kvaasi-Poissoni jaotuse korral on kasutatud assümptootilise normaaljaotuse eeldust.

Haigestumuskordaja standardiseerimine vanuse ja soo järgi



Metoodika:



Lihtne näide magistritöös.



Põhjus:

Rahvastik on vananenud ja inimeste eluiga pikenenud.

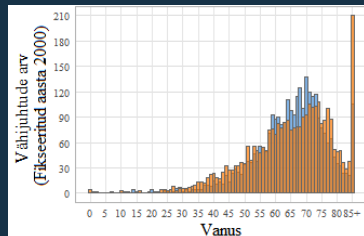
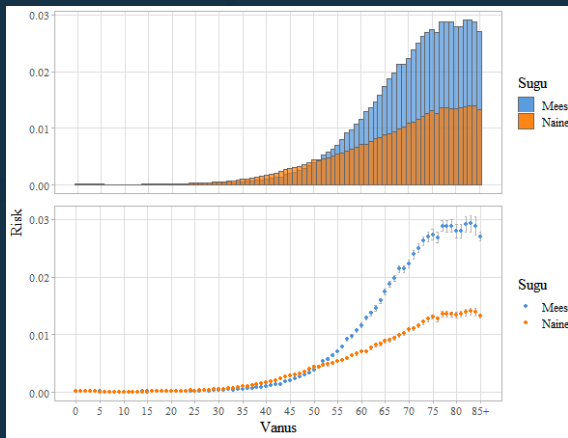


Üldine põhimõte:

Modelleeritud riski järgi on uuritud keskmist haigestunute arvu 10000 inimese kohta, kui vanuseline ja sooline struktuur oleks olnud sama, mis 2016. aastal.

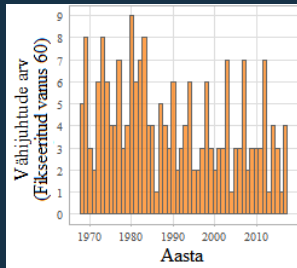
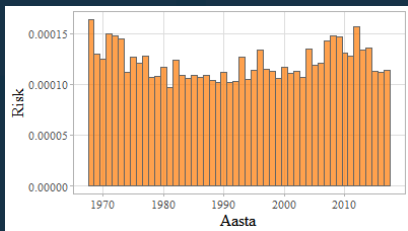
Andmete kirjeldav analüüs: Kõik paikmed

Patsiendi vanuse mõju riskile



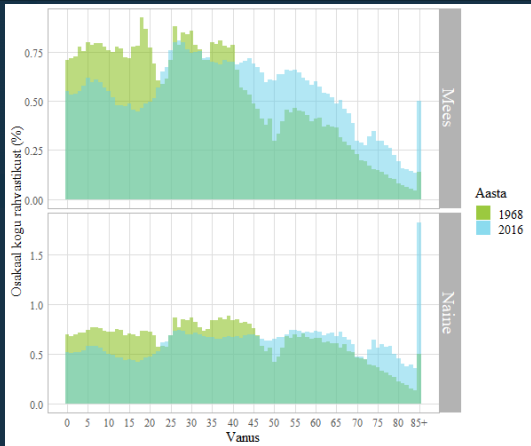
Andmete kirjeldav analüüs: Emakakaelavähk

Diagnoosimise aasta mõju riskile



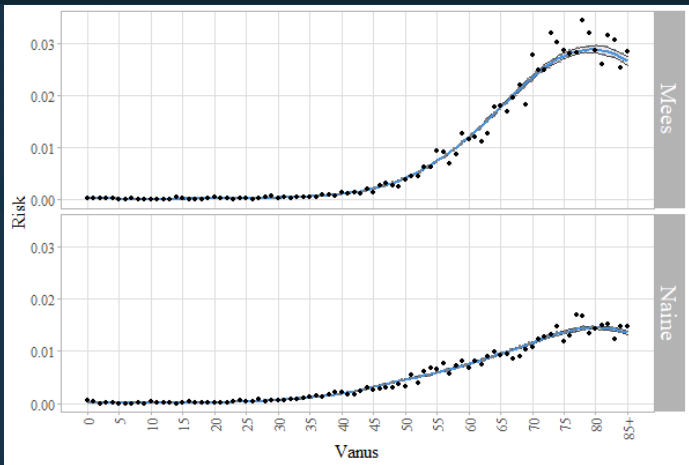
Andmete kirjeldav analüüs:

Rahvastiku sooline ja vanuseline jaotus 1968. ja 2016. aastal

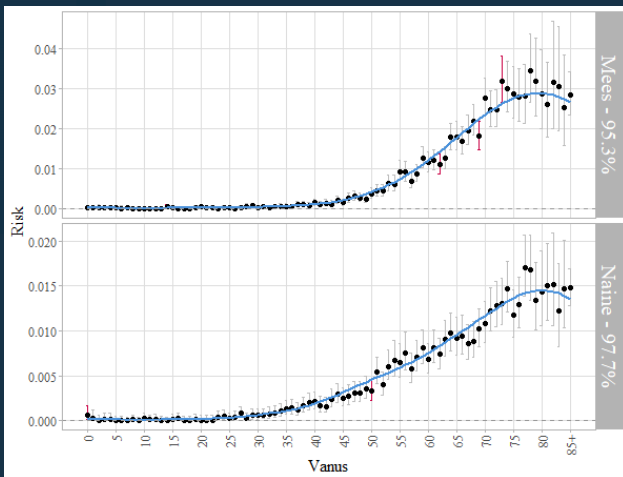


Mudeli efektiivsus

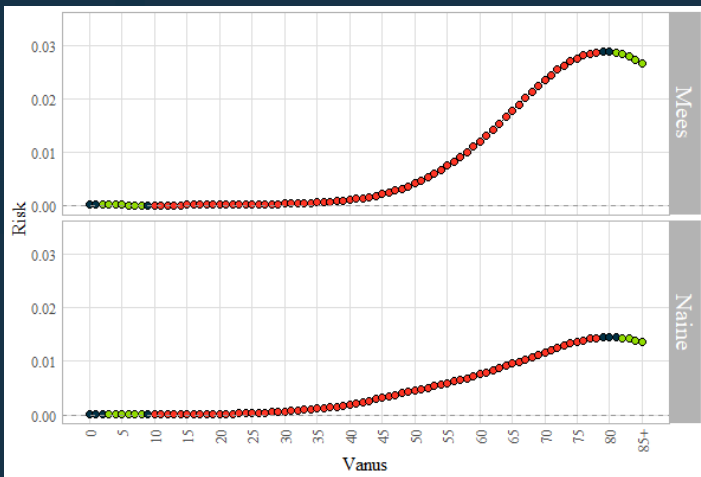
Kõik paikmed: Patsiendi vanuse mõju riskile 2000. aastal



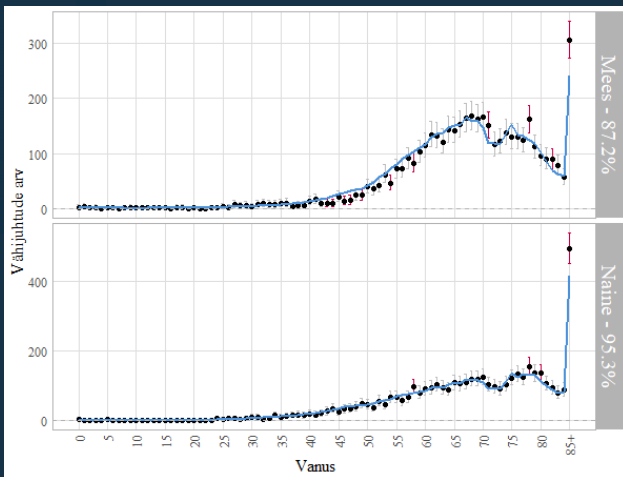
Kõik paikmed: Patsiendi vanuse mõju riskile 2000. aastal



Kõik paikmed: Patsiendi vanuse mõju riskile 2000. aastal

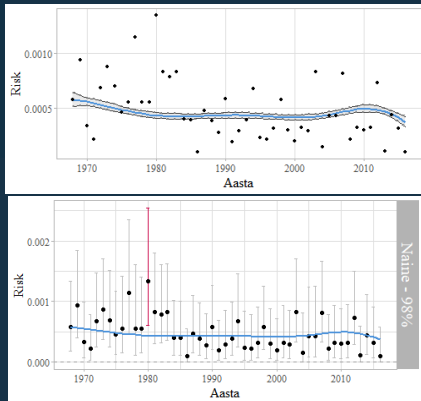


Kõik paikmed: Patsiendi vanuse mõju riskile 2000. aastal

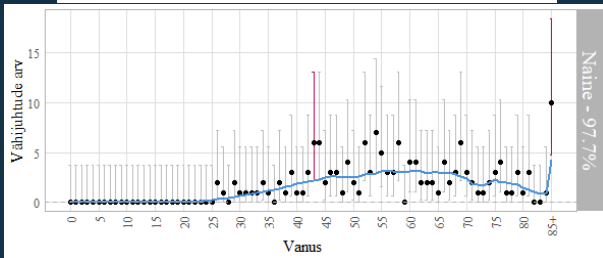
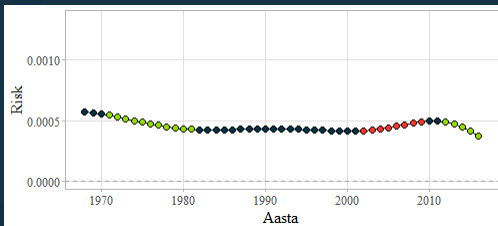


Mudeli efektiivsus

Emakakaelavähk: Diagnoosimise aasta mõju riskile 60aastaste hulgas

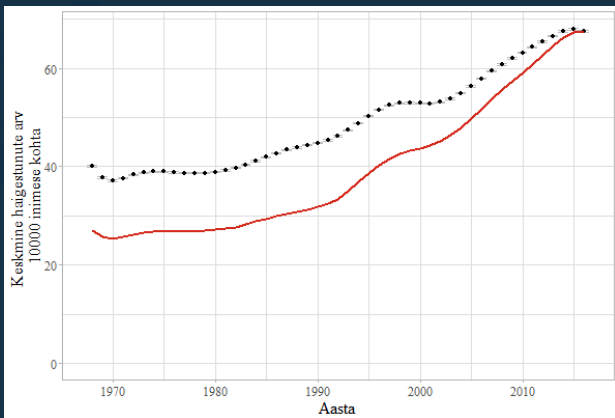


Emakakaelavähk: Diagnoosimise aasta mõju riskile 60aastaste hulgas



Standardiseeritud haigestumuskordaja

Kõikide paikmete korral



Kokkuvõte ja plaanid tulevikus



◆ Antud metoodika põhjal saab väga hästi hinnata vanuse mõju haigestumise riskile



◆ Vähki haigestumise arvu pikemaajaliseks prognoosimiseks kasutada aegridu



◆ Välja arendada erinevad metoodikaid



◆ Luua mudelid ka kõikidele ülejäänutele vähi paikmetele

