

TARTU ÜLIKOOL  
LOODUS- JA TÄPPISTEADUSTE VALDKOND  
MATEMAATIKA JA STATISTIKA INSTITUUT

Hanna-Maria Kukk  
**Väliskeskonna mõju astmariskile**

Matemaatiline statistika  
Bakalaureusetöö (9 EAP)

Juhendajad: PhD Jaanika Kronberg,  
PhD Krista Fischer

TARTU 2024

# VÄLISKESKKONNA MÕJU ASTMARISKILE

Bakalaureusetöö

Hanna-Maria Kukk

## Lühikokkuvõte

Käesoleva bakalaureusetöö eesmärk on uurida geenidoonori kodu ümbritseva väliskeskkonna mõju riskile saada astma, analüüsis arvestatakse ka teadaolevate astma riskifaktoritega. Töös kasutatakse Eesti Geenivaramu andmeid ja EXPANSE projekti raames koostatud keskkonnaandmeid. Statistiline analüüs koosneb Coxi võrdeliste riskide mudelist ja peakomponentanalüüsist. Tulemuste põhjal saab järeldada, et väliskeskkonnal kui komplektil on oluline mõju inimese riskile saada astmadiagnoos. Täpsemalt eksposoom, kus meri on lähedal ja peenosakesi on õhus vähe, vähendab astmariski. Samamoodi on väiksem tõenäosus saada astma, kui elada siseveekogust kaugel ja okaspuude lähedal.

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

**Märksõnad:** astma, eksposoom, elulemusanalüüs, Coxi mudel, peakomponentanalüüs.

# THE INFLUENCE OF THE EXTERNAL ENVIRONMENT ON THE RISK OF ASTHMA

Bachelor thesis

Hanna-Maria Kukk

## Abstract

The aim of this bachelor's thesis is to study the influence of the external environment around gene donors' home on the risk of asthma, the analysis also

takes into account known asthma risk factors. The thesis uses data from the Estonian Genome Centre and the exposome data compiled by the EXPANSE project. Cox proportional hazards model and principal component analysis are used as the main statistical analysis methods. Based on the results, it can be concluded that the external environment as a set has a significant impact on a person's risk of getting a diagnosis of asthma. Specifically, asthma risk is lower in an environment which is close to sea and has few fine particles in the air. Similarly, person is less likely to get asthma if the home is far from inland water bodies and near conifers.

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

**Key Words:** asthma, exposome, survival analysis, Cox model, principal component analysis.

# Sisukord

<b>Sissejuhatus</b>	<b>5</b>
<b>1 Astma</b>	<b>6</b>
<b>2 Eksposoom</b>	<b>8</b>
2.1 Õhusaaste . . . . .	8
2.2 CORINE'i rohealad . . . . .	9
2.3 Roheluse indeks . . . . .	11
<b>3 Metoodika</b>	<b>12</b>
3.1 Elulemusandmete analüüs . . . . .	12
3.2 Coxi võrdeliste riskide mudel . . . . .	13
3.3 Tunnuste väärtuste standardiseerimine . . . . .	16
3.4 Peakomponentanalüüs . . . . .	17
<b>4 Andmete analüüs</b>	<b>19</b>
4.1 Üldtunnuste valik . . . . .	19
4.2 Eksposoomitunnuste valik . . . . .	20
4.3 Valim . . . . .	23
4.4 Coxi mudel . . . . .	24
4.5 Peakomponentanalüüs . . . . .	28
<b>5 Tulemuste arutelu</b>	<b>34</b>
<b>Kokkuvõte</b>	<b>37</b>

<b>Kasutatud allikad</b>	<b>39</b>
<b>Lisad</b>	<b>44</b>
Lisa 1 . . . . .	45
Lisa 2 . . . . .	46
Lisa 3 . . . . .	48
Lisa 4 . . . . .	50

## Sissejuhatus

Töös on kasutatud Tartu Ülikooli Eesti Geenivaramu, Eesti Haigekassa Ravikindlustuse andmekogu ja EXPANSE projekti poolt koostatud keskkonnaandmeid. Üle-Euroopalise eksposoomi projekti EXPANSE eesmärk on luua eksposoomiandmestik ning tööriistad, et analüüsida elukeskkonna mõju inimeste tervisele [1] ning selles töös uuritakse eksposoomi ja tulevikus tekkiva astma seoseid.

Töö algab teoreetilise osaga, mis jaotub kolmeks peatükiks. Esimeses peatükis on kirjeldatud lühidalt olulisemaid teadmisi astmast. Teisest peatükist leiab kirjelduse EXPANSE projekti käigus loodud eksposoomiandmestikule ning osade eksposoomitunnuste varasemate uuringutega kirjeldatud seosed astmariskiga. Kolmandas peatükis on antud ülevaade analüüsis kasutatud metoodikast.

Töö analüüsi osa jaotub viieks: üldtunnuste ja eksposoomitunnuste valiku selgitus, uuritava valimi moodustumine, elulemusanalüüsis Coxi mudeli koostamine ja eksposoomitunnuste peakomponentanalüüs. Elulemusanalüüsis on ajaskaalaks aeg astma diagnoosi saamiseni kümne aasta jooksul alates 01.01.2013. Arvestatud on paremalt tsenseeritud andmetega. Töös on koostatud Coxi võrdeliste riskide mudelid, kus argumenttunnustena on kasutatud nii keskkonnatunnuseid kui ka peakomponente. Peakomponendid on leitud üle keskkonnatunnuste väärtuste. Analüüsi läbiviimisel ja jooniste koostamisel on kasutatud rakendustarkvara R. Töö on kirjutatud programmiga L<sup>A</sup>T<sub>E</sub>X.

Töö autor tänab juhendajaid Krista Fischerit ja Jaanika Kronbergi heade nõuanete ja selgituste eest ning TÜ Eesti Geenivaramut andmete kasutamise loa eest.

# 1 Astma

Rahvusvahelise Haiguste Klassifikatsiooni 10. versiooni ehk RHK-10 põhjal on astma (RHK-10 klassifikaatori koodiga J45) alumiste hingamisteede krooniline haigus, mis jaotub neljaks: valdavalt allergiline astma, mitteallergiline astma, segatüüpi astma ja täpsustamata astma. Samuti on eraldi olemas astmaatilise seisundi haigus (koodiga J46), mis jaotub kaheks: astmaatiline seisund ja raskekujuline äge astma. [2]

Antud töös erinevaid diagnoose omavahel ei võrrelda ning korraga vaadeldakse diagnoose koodiga J45 ja J46. Olgu edaspidi astmaatik inimene, kellel esineb andmebaasis

- vähemalt kahel korral diagnoos koodiga J45 või J46;
- või üks diagnoos koodiga J45 või J46 ning lisaks üks ravimiretsept, mis on seotud J45 või J46 diagnoosiga.

Astma diagnoosi saanud on enamasti kogenud episoodiliselt tekkivaid hingamisraskusi, õhupuudust, raskustunnet rindkeres ja hootist köha. Osadel astmaatikutel esineb ka kiuneid või vilinaid rinnus. Raviga on võimalik astma kontrolli alla saada nii, et eelnevaid sümptomeid enam ei esine ning kehalises tegevuses pole piiranguid. Kahjuks on astma ravitulemused halvemad ärevushäire või depressiooniga patsientidel. Eestis on ligikaudu igal viieteistkümnendal inimesel diagnoositud astma, kuid kuna astma on siiski veel vähe diagnoositud, siis selle sümptomitega inimeste arv võib olla veelgi suurem. [3]

Ägeda astmahoo tõttu võib ka surra, näiteks Inglismaa kaotab igal aastal astma pärast umbes 1200 inimest. Surmajuhtude vähendamiseks on seal analüüsitud erinevaid riskitegureid ning on leitud, et patsiendi keskkond on oluline mõjutav tegur. Astmaatikutel on soovitatav vältida tubakasuitsu, kuna see on üks suurimaid hingamisteede ärritajaid. [4]

Astma teadaolevateks riskifaktoriteks on kokkupuude allergeenidega ja hingamisteede ärritajatega, nagu tolm, loomakarvad, hallitus. Astmaatikute keha tekitab limaskestas põletikku, et kaitsta end väliskeskkonna ärritajate vastu. Neil, kellel on lapsena diagnoositud atoopiline dermatiit ehk nahapõletik või allergiline nohu, on täiskasvanueas suurem risk haigestuda astmasse. Samuti soodustavad astma saamist kõrgvererõhktõbi, südamelihase infarkt ja südamepuudulikkus. [3]

Eesti Kopsuliit soovitab astmaatikutel ka vältida õietolmu või külma õhu sissehingamist [5]. Maailma Terviseorganisatsiooni kohaselt on suurendanud astma levimust linnastumine ja kokkupuude õhusaastega juba varases eas [6].



## 2 Eksposoom

EXPANSE projekti käigus on aastatest 2000 kuni 2020 mudeldatud erinevate Euroopa riikide väliskeskkonna andmeid. Füüsilise keskkonna tingimuste kohta, nagu veekogude kaugus, roheluse, hallide alade (nt asfalt, ehitised) või öise tehisvalguse hulk, on kogutud info satelliidipiltide põhjal iga 1-5 aasta tagant. Igale elupaigale on arvutatud erinevaid elukvaliteeti või linnasust määravaid indekseid. Õhukvaliteedist on iga-aastaseid andmeid nii peenosakeste kui ka osooni või lämmastikdioksiidi kohta. Samuti on igal piirkonnal mõõtmistulemuste põhjal mudeldatud igapäevased keskmised, miinimum- ja maksimumtemperatuurid kakskümmend aastat järjest. Andmete põhjal on loodud eksposoomikaardid, mida saab kasutada erinevates uuringutes koduste aadresside või läbitavate teekondade eksposoomi uurimiseks. Keskkonnaandmed on lingitud Eesti geenivaramu biopanga andmetega nii, et teadlastel ei ole aadressidele ega koordinaatidele ligipääsu. [7]

### 2.1 Õhusaaste

Sõiduvahendi kütuse põlemisel eraldub õhku lämmastikdioksiid ( $NO_2$ ), mis ärritab hingamisteid ning soodustab astma teket. Õhus olles  $NO_2$  reageerib teiste kemikaalidega ja moodustab nii peenosakesi kui ka osooni. [8]

Eestis on lämmastikdioksiidi aasta keskmine lubatud maksimaalne kogus  $40 \mu g/m^3$ , mille ületamisel rakendatakse taset vähendavaid abinõusid [9]. Maailma Terviseorganisatsioon (WHO) ei soovita üle  $10 \mu g/m^3$  aastast keskmist taset ning lühiajalise 24 tunni keskmine kokkupuude ei tohiks olla üle  $25 \mu g/m^3$  [10].

Maapinna lähedal olev osoon on ohtlik õhusaaste astmaatikutele. Seda on enamasti palju linnas soojadel päikeselistel päevadel, kuid see kandub lihtsasti tuulega edasi ka maapiirkondadesse. [11]

WHO soovitusel osooni keskmine 8 tunni tase ei tohiks ületada  $100 \mu g/m^3$  ning kuue kuu keskmine päeva kogus ei tohiks ületada  $60 \mu g/m^3$  [10]. Osooni kõrgeim

8 tunni keskmise lubatud tase on Eestis  $120 \mu\text{g}/\text{m}^3$  [9].

Väikeste osakeste suur hulk õhus on potentsiaalselt ohtlik astmaatikutele, sest need jõuavad kopsudesse ja sealt edasi ka vereringesse. Osakesed, mille diameeter on alla 10 mikromeetri, kvalifitseeritakse peenete osakeste ( $PM_{10}$ ) alla ning osakesed, mille diameeter on alla 2,5 mikromeetri, on ülipeened ( $PM_{2,5}$ ). Mitmed varasemad uuringud on leidnud, et  $PM_{10}$  ja  $PM_{2,5}$  suur sisaldus õhus halvendab kopsu normaalset tööd ning soodustab astma teket ja erinevate kopsuhaiguste diagnoosiga inimestel ka varajast surma. Peenosakesed halvendavad ka ülejäänud keskkonna seisuga, sest nende maha sadestudes väheneb mulla toitainete rikkus või muutub erinevate veekogude biokeemiline tasakaal. [12]

Eestis on lubatud  $PM_{10}$  maksimaalne kogus  $40 \mu\text{g}/\text{m}^3$  aasta keskmise väärtuse jaoks, kuid  $PM_{2,5}$  piir on  $25 \mu\text{g}/\text{m}^3$  [9]. Maailma Terviseorganisatsiooni soovituslikud piirid on rangemad: aasta keskmise  $PM_{10}$  maksimaalne piir on  $15 \mu\text{g}/\text{m}^3$  ja  $PM_{2,5}$  oma on  $5 \mu\text{g}/\text{m}^3$  [10].

Eestis on kõige halvemad ülipeenete osakeste näitajad õhus kuues omavalitsuses: Tartu ja Võru linnas, Luunja, Kastre, Valga ja Antsla vallas. Nendes kohtades  $PM_{2,5}$  hulk küündib  $8,5 \dots 10 \mu\text{g}/\text{m}^3$  keskmiselt aastas. Üheski Eesti omavalitsuses ei ole aasta keskmine näitaja alla WHO soovitatava piiri ehk alla  $5 \mu\text{g}/\text{m}^3$ . [13]

## 2.2 CORINE'i rohealad

CORINE'i maakatte andmekogum sisaldab ühtlast ja üksikasjalikku infot maakasutuse kohta üle Euroopa. Maa on jagatud 44 erinevasse klassi ning ühe alaühiku suurus on minimaalselt 25 hektarit, kuid tuvastatakse ka minimaalselt 100 meetri laiused lineaarsed objektid, nagu teed ja jõed. Andmekogumit uuendatakse iga kuue aasta tagant ning sellega on liitunud praeguseks 39 erinevat riiki. Infot kogutakse satelliitidega, mille täpsus viimase korje ajal, aastal 2018 oli 10 meetrit ning 85% aladest määrati temaatiliselt õigesse klassi. [14]

Copernicuse näidiskaardist [14], kust näeb CORINE'i rohealade paiknemist, on võetud välja kuvatõmmis Tartu linnast. Lisas 1 on kuvatõmmisega moodustatud Tartu linna CORINE'i alade kaart, koos alade koodide ja tähenduse legendiga. Tartu linnas on CORINE'i rohealad piirkonnad, mis on järgnevate koodide värvidega: 141, 142, 211, 231, 242, 243, 312, 313. Ülejäänud koodidega alad ei ole CORINE'i rohealad. Kaardile on musta tekstiga juurde kirjutatud rohealade (või roheala kõrval asuvate) piirkondade üldtuntud nimetused, mis aitavad kaardi vaatajal paremini orienteeruda. Tartu linnas on ulatuslikult CORINE'i rohealad Ihaste, Tähtvere ja Kvissentali linnaosades. Raadi-Kruusamäe linnaosa piirid lõikavad CORINE'i rohealad nagu Raadi kalmistu, Raadi park ja Raadi looduskaitseala. Keslinna linnaosas on CORINE'i rohealaks Toomemäe park. Annelinna ääres on CORINE'i rohealadeks Anne kanali ja Emajõe vaheline linna roheala, Linnupargi, Tartu Tammik ja Anne looduskaitseala. Veeriku, Ränilinna ja Maarjamõisa linnaosad ääristavad CORINE'i rohealad, mis on erineva sihtotstarbega maa klassi all. Need on pigem tüher- või rohumaad suurte magistraalide kõrval. Ülejõe ja Supilinna linnaosade piiril on väike osa CORINE'i roheala Emajõe kaldal. CORINE'i rohealad puuduvad Jaamamõisa, Karlova, Ropka, Ropka tööstuse, Variku, Tammelinna ja Vaksali linnaosades. Kaardilt on näha, et mitmed linnapargid või erakinnistute aiad on välja jäänud CORINE'i rohealadest, sest need on liiga väikse pindalaga, kuigi ka nendel aladel on potentsiaalne mõju astmaatikutele. Tulemuste tõlgendamisel peab ka arvestama, et Tartu linna näitel on CORINE'i rohealad väga eriilmelised, näiteks metsasalud Ihastes või suurte magistraalide kõrval olevad rohualad Veerikus.

Linnaelanikud peaksid WHO soovitusel elama vähemalt 300 meetri ehk umbes 5 minutilise jalutuskäigu kaugusel vabalt ligipääsetavast rohealast. EXPANSE projektis defineeritakse CORINE'i rohealana 20 klassi 40-st, mille all on nii pargid, põllumaad, istikud, karjamaad, metsad, võsad kui ka rohtlad. [15]

## 2.3 Roheluse indeks

Väliskeskkonna rohelist ilmetab EXPANSE projektis ka satelliitide abil mõõdetud NDV (*Normalized Difference Vegetation*) indeks. Antud indeksi väärtused arvutatakse  $6,25 \text{ m}^2$  täpsusega infrapuna ja punase valguse erinevuste põhjal, mida fotosünteesivad taimed peegeldavad või neelavad. EXPANSE projektis vastavad väärtused on nulli ja ühe vahel, kus madalad väärtused näitavad hõredat taimeistikku ning kõrged tihedat. Siin veekogud on NDV indeksist välja jäetud. EXPANSE projektis ilmetab rohelist veel ka MSAV indeks ehk MVI. Selles indeksis arvestatakse arvutamisel maapinna heleduse või tumedusega ehk see on sobivam NDV indeksist avatud mullapindadega aladel või tuvastab paremini noort taimeistikku. Selle väärtused on samamoodi vahemikus nullist üheni. [15]

Elukeskkonna roheline parandab astmaatikute tervist mitmel viisil: vähendab stressi, suurendab kehalist liikuvust, puud neelavad enesesse õhusaastet. Kuid roheline võib ka halvendada astmaatikute seis: õietolm ärritab astmaatikute hingamisteid ning tihe kõrge taimeistik takistab tuule levikut ehk suurendab õhusaaste kontsentratsiooni. Sellest tulenevalt on rohealad osutunud varasemates uuringutes nii positiivseks kui negatiivseks teguriks astmale. [16]

## 3 Metoodika

Siinse peatüki järgnevas 3.1 ja 3.2 alapeatükkides kirjeldatav metoodika põhineb raamatu *Statistical methods for survival data analysis* (2013) lehekülgedel 1-12, 161-163 ja 282-288 [17].

### 3.1 Elulemusandmete analüüs

Elulemusandmete analüüsi kasutatakse tihti meditsiinis, et uurida aega kindla sündmuseni, näiteks aega surmani, haigestumiseni või tervenemiseni. Võttes andmete hulka ka muud faktortunnused, näiteks harjumuste või füüsiliste karakteristikute kohta, saab analüüsida riskitegurite olemasolu ja nende mõju suurust valitud sündmusele.

Olgu indiviidide arv  $n \in \mathbb{N}$ , keda vaadeldakse fikseeritud algusajast  $t_0$  fikseeritud lõppajani  $t_*$ . Tähistagu  $A$  huvipakkuvat sündmust. Iga indiviidi  $i$  kohta fikseeritakse sündmuse  $A$  toimumise aeg  $t_i$ , kus  $i = 1, \dots, n$ . Kui sündmuse toimumise aeg ei lange vaatlusperioodi sisse ehk  $t_i > t_*$  või  $t_i < t_0$ , siis on vaatlus  $t_i$  tsenseeritud. Näiteks osad indiviidid pole uuringu lõpuks veel haigestunud, kuid võivad haigestuda hiljem. Vaatlus  $t_i$  on tsenseeritud ka siis, kui puudub info  $t_i$  kindlalt määratlemiseks, näiteks indiviid on enne vaatlusaja lõppu uuringust lahkunud. Uuringutes on kõige tavapärasem, et osad vaatlused on paremalt tsenseeritud. Öeldakse, et vaatlus  $t_i$  on paremalt tsenseeritud, kui  $t_i > t_*$  või kui määratlemata  $t_i$  kohta on teada fikseeritud aeg  $t_a \leq t_i$ .

Olgu fikseeritud üks vaatlusalune, kelle puhul sündmuse  $A$  toimumise aeg on  $T \in [t_0; t_*]$  ja olgu fikseeritud juhuslik aeg  $t$ .  $T$  on pidev juhuslik suurus. Erinevate eluliste andmetega tegeledes peab arvestama, et aja  $T$  jaotus ei vasta enamasti ühelegi tuntud parameetrilisele jaotusele ja seda tuleb kirjeldada mitteparameetriliselt. Muutuja  $T$  jaotust saab iseloomustada üleelamisfunktsiooniga  $S(t)$ , tihedusfunktsiooniga  $f(t)$  ja riskifunktsiooniga  $h(t)$ .

Üleelamisfunktsioon  $S(t)$  näitab tõenäosust, et vaatlusalusel ei toimu sündmust  $A$  enne aega  $t$ .  $S(t)$  on sündmuse  $T > t$  tõenäosus. Üleelamisfunktsioon  $S(t)$  on defineeritud kui:

$$S(t) = P(T > t) = 1 - F(t),$$

kus  $F(t)$  on juhusliku suuruse  $T$  jaotusfunktsioon.

Riskifunktsioon  $h(t)$  on defineeritud kui:

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

kus  $\Delta t$  on ajahetk vahetult pärast aega  $t$ .

Kehtib seos:

$$h(t) = \frac{f(t)}{S(t)}.$$

Riskifunktsioon  $h(t)$  iseloomustab tõenäosust, et sündmus  $A$  toimub vahetult pärast hetke  $t$  tingimisel, et hetkeks  $t$  sündmust veel toimunud ei ole.

Oletame, et saab eeldada, et riskifunktsioon on ligikaudu konstantne vahemikus  $(t, t + \Delta t)$  mingi väikese ajaühiku  $\Delta t$  korral. Kui andmestikus puuduvad tsenseeritud vaatlused ja leiduvad  $m \in \mathbb{N}$  vaatlust, kus  $t_j > t$ ,  $j = 1, \dots, m \leq n$ , ning nende  $m$  vaatluse hulgas leidub omakorda  $l$  vaatlust, kus  $t_u \in (t, t + \Delta t)$ ,  $u = 1, \dots, l \leq m$ , siis riskifunktsiooni hinnang on leitav valemiga:

$$\hat{h}(t) = \frac{l}{m\Delta t}.$$

Tsenseeritud vaatlustega ei saa kasutada eelnevat valemit.

## 3.2 Coxi võrdeliste riskide mudel

Coxi võrdeliste riskide mudel on poolparameetriline meetod, mis ei eelda parameetrilist jaotust elukestusele  $T$ .

Oletame, et sündmuse  $A$  toimumise tõenäosust indiviidil  $j$  mõjutavad  $p$  erinevat riskitegurit  $x_{ij}$ ,  $i = 1, \dots, p$ . Vaatleme kahte indiviidi riskitegurite väärtustega  $x_1 = (x_{11}, x_{21}, \dots, x_{p1})$  ja  $x_2 = (x_{12}, x_{22}, \dots, x_{p2})$ . Neile vastavad riskifunktsioonid on  $h(t|x_1)$  ja  $h(t|x_2)$ . Võrdeliste riskide eeldus tähendab seda, et kahe suvalise vaatluse riskifunktsiooni suhe peab olema konstantne ajas:

$$\frac{h(t|x_1)}{h(t|x_2)} = \text{const} \quad \forall t \in [t_0, t_*].$$

Sellest eeldusest tulenevalt iga suvalise  $x$  korral riskifunktsioon  $h(t|x)$  sõltub ajas muutuvast riskifunktsioonist  $h_0(t)$  ja riskitegurite väärtustest sõltuvast funktsioonist  $g(x_1, x_2, \dots, x_p)$ . Eelneva kahe riskifunktsiooni suhte saab siis ümber kirjutada järgmiselt:

$$\frac{h(t|x_1)}{h(t|x_2)} = \frac{h_0(t)g(x_{11}, x_{21}, \dots, x_{p1})}{h_0(t)g(x_{12}, x_{22}, \dots, x_{p2})} = \frac{g(x_1)}{g(x_2)} \quad \forall t \in [t_0, t_*].$$

Tavapärane on kasutada funktsioonina  $g(x)$  eksponentfunktsiooni, mille korral Coxi võrdeliste riskide mudel avaldub kui:

$$h(t|x) = h_0(t)e^{x_1\beta_1 + \dots + x_p\beta_p} \quad \forall t \in [t_0, t_*].$$

Olgu kahel indiviidil ühe riskiteguri väärtused erinevad:  $x_{11} \neq x_{12}$ , kuid kõik teised riskitegurite väärtused võrdsed:  $x_{i1} = x_{i2} \quad \forall i = 2, \dots, p$ . Kahe vaatluse riskifunktsioonide suhe on siis

$$\frac{h(t|x_1)}{h(t|x_2)} = \frac{h_0(t)e^{x_{11}\beta_1 + \dots + x_{p1}\beta_p}}{h_0(t)e^{x_{12}\beta_1 + \dots + x_{p2}\beta_p}} = \frac{h_0(t)e^{x_{11}\beta_1} \dots e^{x_{p1}\beta_p}}{h_0(t)e^{x_{12}\beta_1} \dots e^{x_{p2}\beta_p}} = \frac{e^{x_{11}\beta_1}}{e^{x_{12}\beta_1}} = e^{(x_{11} - x_{12})\beta_1}.$$

Kui riskiteguri väärtus  $x_{11}$  erineb riskiteguri väärtusest  $x_{12}$  ühe ühiku võrra, siis riskifunktsioonide suhe on võrdne  $e^{\beta_1}$ . See tähendab, et kahe indiviidi risk sündmuse  $A$  toimumiseks erineb  $e^{\beta_1}$  korda. Kui  $e^{\beta_1} > 0$ , siis antud riskitegur suurendab sündmuse  $A$  toimumise riski. Kui  $e^{\beta_1} < 0$ , siis antud riskitegur vähendab sündmuse

$A$  toimumise riski.

Riskiteguri kordajate komplekti hinnangu  $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$  leidmisel saaks iga riskiteguri hinnanguga  $\hat{\beta}_i$ ,  $i = 1, \dots, p$  interpreteerida riskiteguri mõju sündmusele  $A$ . Samuti on hinnatav siis iga indiviidi riskitegurite komplekti  $x_i = (x_{1i}, x_{2i}, \dots, x_{pi})$ ,  $i = 1, \dots, n$ , mõju sündmuse  $A$  toimumisele ning on võrreldav kahe juhusliku indiviidi sündmuse  $A$  toimumise riskide erinevus.

Kui sündmuse  $A$  toimumisaeg  $T$  on teadaoleva parameetrilise jaotusega, siis on  $\hat{\beta}$  leitav suurima tõepära meetodiga. Kui me seda eeldust teha ei saa, siis tavapärase tõepärafunktsioon ei toimi, sest Coxi mudelis on igal vaatlusel sama riskifunktsioon  $h_0(t)$ , mille kuju ei ole teada. Sellest tulenevalt Coxi mudeliga ka ei prognoosita sündmuse  $A$  toimumist läbi riskifunktsiooni, vaid hinnatakse riskitegurite mõju sündmusele läbi kordajate hinnangute  $\hat{\beta}_i$ ,  $i = 1, \dots, p$ . Parameetrite  $\beta_i$  hindamine toimub erinevalt vastavalt sellele, kas esineb võrdseid  $t_i$ ,  $i = 1, \dots, n$ , väärtuseid valimis või mitte. Võrdsete väärtuste olemasolul sõltub hinnangu arvutamise valem ka sellest, kas aega vaadeldakse diskreetse või pideva tunnuseks. Järgnevalt on kirjeldatud lihtsam juht ehk kui võrdseid väärtuseid ei esine, kuid arvestatakse paremalt tsenseeritusega.

Olgu  $n - k$  indiviidi, kelle sündmuse toimumise ajad  $t_i^+$ ,  $i = 1, \dots, n - k$ , on paremalt tsenseeritud, ning  $k$  mittetsenseeritud vaatlust  $t_i$ ,  $i = 1, \dots, k$ . Kokku on  $n$  vaatlust. Teadaolevad  $k$  aega tähistatakse vastavalt väärtuste järjestusele  $t_{(1)} < t_{(2)} < \dots < t_{(k)}$  ning vastavate indiviidide riskitegurite komplektid tähistatakse samasuguste indeksitega  $x_{(1)}, x_{(2)}, \dots, x_{(k)}$ . Iga tsenseeritud  $n - k$  aja kohta on teada tsenseerimisaeg  $t_{ai}$ , kus  $t_{ai} \leq t_i^+$ ,  $i = 1, \dots, n - k$  ja  $t_{ai} \leq t^*$ , ning olgu tsenseerimisaegad tähistatud vastavalt nende järjestatud väärtustele  $t_{(a1)} < t_{(a2)} < \dots < t_{(a(n-k))}$ . Tsenseeritud vaatluste riskitegurite komplektid tähistatakse samasuguste indeksitega  $x_{(a1)}, x_{(a2)}, \dots, x_{(a(n-k))}$ . Olgu fikseeritud juhuslik aeg  $t_{(i)}$ , kus  $i \in \{1, \dots, k\}$ , ja defineeritud vaatluste indeksite rühm  $R(t_{(i)})$ , kellel on võimalus sündmuse  $A$  toimumiseks ajal  $t_{(i)}$ . Tsenseeritud aegadega vaatluste jaoks see tä-



hendab, et  $aj \in R(t_{(i)})$ , kui  $t_{(aj)} \geq t_{(i)}$ ,  $j = 1, \dots, n-k$ . Mittetsenseeritud aegade vaatluste puhul  $j \in R(t_{(i)})$ , kui  $t_{(j)} \geq t_{(i)}$ ,  $j = 1, \dots, k$ . Olgu üks juhuslik indiviid  $i$ , nii et  $i \in R(t_{(i)})$ , ja tema riskitegurite komplekt  $(x_{1(i)}, \dots, x_{p(i)})$ . Tõenäosus, et sündmus  $A$  toimub ajal  $t_{(i)}$  on arvatav valemiga

$$\frac{e^{x_{1(i)}\beta_1 + \dots + x_{p(i)}\beta_p}}{\sum_{l \in R(t_{(i)})} e^{x_{1(l)}\beta_1 + \dots + x_{p(l)}\beta_p}} = \frac{e^{x_{(i)}\beta}}{\sum_{l \in R(t_{(i)})} e^{x_{(l)}\beta}}.$$

Järgnevalt saab arvutada eelneva tõenäosuse igale teadaoleva sündmuse toimumise ajale  $t_{(i)}$ ,  $i = 1, \dots, k$ :

$$L(\beta) = \prod_{i=1}^k \frac{e^{x_{1(i)}\beta_1 + \dots + x_{p(i)}\beta_p}}{\sum_{l \in R(t_{(i)})} e^{x_{1(l)}\beta_1 + \dots + x_{p(l)}\beta_p}} = \prod_{i=1}^k \frac{e^{x_{(i)}\beta}}{\sum_{l \in R(t_{(i)})} e^{x_{(l)}\beta}},$$

mis on tõepärafunktsioon riskitegurite kordajate komplektile  $\beta$ . Selline  $\beta$  väärtus, mille korral tõepärafunktsioon  $L(\beta)$  on maksimaalne, ongi (suurima tõepära) hinnang  $\hat{\beta}$ . Hinnangu kordajate komplekti  $(\hat{\beta}_1, \dots, \hat{\beta}_p)$  leidmiseks on vaja lahendada  $p$  erinevat võrrandit:

$$\frac{\partial(\log L(\beta))}{\partial \beta} = 0.$$

Edasises analüüsi kirjelduses kasutan Coxi mudeli matemaatilise kuju  $h(t|x) = h_0(t)e^{x_1\beta_1 + \dots + x_p\beta_p}$  asemel lihtsamat kirjalpilti:

$$(t, A) \sim x_1 + x_2 + \dots + x_p.$$

### 3.3 Tunnuste väärtuste standardiseerimine

Skaleerimine ehk standardiseerimine muudab tunnuse  $x$  algsed väärtused  $x_1, \dots, x_n$  selliseks, et uus standardväärtus  $x_{si}$ ,  $i = 1, \dots, n$ , kirjeldab mitme standardhälbe

$\sigma$  kaugusel on algväärtus  $x_i$  tunnuse keskväärtusest  $\mu$ :

$$x_{si} = \frac{(x_i - \mu)}{\sigma}.$$

Tunnuse  $x$  standardhälve  $\sigma$  ja keskväärtus  $\mu$  hinnatakse vaatluste väärtuste põhjal. Standardväärtus  $x_{si}$  on positiivne, kui selle algväärtus on üle keskmise, ning negatiivne, kui algväärtus on alla keskmise. [18]

Kasutades skaleeritud väärtusi Coxi mudelis saab korraga analüüsida tunnuseid, mille väärtuste skaalad on väga erinevad. Coxi mudelis on skaleeritud väärtustega tunnusel  $e^\beta$  kordne riskide erinevus kahel juhuslikul vaatlusel, kui nende vaatluste standardväärtused erinevad teineteisest ühe ühiku võrra ehk algväärtused erinevad ühe standardhälbe võrra.

### 3.4 Peakomponentanalüüs

Siinse alapeatüki peakomponentanalüüsi meetodika on kirjeldatud raamatu *Applied Multivariate Statistical Analysis* (1992) lehekülgede 357-362 põhjal [19].

Mitmete omavahel erinevalt korreleeritud tunnuste olemasolul saab peakomponentanalüüsiga lihtsamini ja kokkuvõtlikumalt arvestada tunnuste korrelatsioonimaatriksis esinevate mustritega.

Olgu analüüsis  $p$  erinevat tunnust (või riskitegurit)  $x_1, \dots, x_p$ , ja nende kovariatsioonimaatriks  $\Sigma$ . Tähistagu  $X$  tunnuste vektorit  $X = [x_1, \dots, x_p]$ .

Tunnustest  $x_i$  saab moodustada lineaarkombinatsioone  $Y_i = l_{1i}^T x_1 + l_{2i}^T x_2 + \dots + l_{pi}^T x_p = l_i^T X$ , kus  $l_{ji} \in \mathbb{R}$  on suvalised kordajad ja  $i = 1, \dots, q \in \mathbb{N}$ ,  $j = 1, \dots, p$ . Lineaarkombinatsiooni  $Y_i$  dispersiooni ja kahe juhusliku lineaarkombinatsiooni  $Y_i$  ja  $Y_k$ , kus  $i, k = 1, \dots, q \in \mathbb{N}$ , kovariatsiooni valemid on

$$D(Y_i) = D(l_i^T X) = l_i^T \Sigma l_i,$$

$$\text{Cov}(Y_i, Y_k) = l_i^T \Sigma l_k.$$

Peakomponendid on sellised  $p$  erinevat lineaarkombinatsiooni, mis on teineteisest sõltumatud ehk  $\text{Cov}(Y_i, Y_k) = 0$ , kus  $i, k = 1, \dots, p$ , ning mille kordajad  $l_{ji}$ , kus  $j, i = 1, \dots, p$ , valitakse nii, et hajuvus oleks võimalikult suur. Hajuvuse (dispersiooni) maksimeerimisel arvestatakse, et kordajate vektori  $l_i$  pikkus on üks:  $l_i^T l_i = 1$ ,  $i = 1, \dots, p$ . Esimene peakomponent on lineaarkombinatsioon  $Y_1 = l_1^T X$ , teine peakomponent on  $Y_2 = l_2^T X$ ,  $\dots$ ,  $p$ . peakomponent on  $Y_p = l_p^T X$ .

Peakomponendid on lihtsasti leitavad kovariatsioonimaatriksi omaväärtuste ja omavektorite kaudu. Olgu kovariatsioonimaatriksi  $\Sigma$  omaväärtused  $\lambda_1, \dots, \lambda_p$ , kus  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ . Olgu omaväärtustele vastavad omavektorid  $e_1, \dots, e_p$  ning vastavad paarid  $(\lambda_1, e_1), \dots, (\lambda_p, e_p)$ . Nende kaudu avaldub  $i$ . peakomponent, kus  $i = 1, \dots, p$ :

$$Y_i = e_{1i} X_1 + \dots + e_{pi} X_p = e_i^T X.$$

Peakomponendid on leitavad ka korrelatsioonimaatriksi  $P$  põhjal, kuid selle peakomponendid ei ühti kovariatsioonimaatriksi  $\Sigma$  põhjal leitavate peakomponentidega.

Kõigi  $n$  vaatluse  $p$  tunnuse väärtuste hajuvuse kirjeldamiseks on vaja kõiki  $p$  peakomponenti. Kui mitmed tunnused on omavahel korreleeritud, siis piisab suurema osa hajuvuse kirjeldamisest ka ainult mõnest peakomponendist.

Kui analüüsitavate tunnuste väärtuste skaalad on väga erinevad, siis on soovitatav rakendada peakomponentanalüüsi standardiseeritud väärtustega tunnustele. Kui kõik  $p$  tunnust on standardiseeritud, siis tunnuste kovariatsiooni- ja korrelatsioonimaatriks on võrdsed. Standardiseeritud tunnustele hinnatud peakomponendid kirjeldavad tunnuste hajuvust nii, et tunnused on võrdsema kaaluga peakomponendile ehk suurte väärtuste vahemikuga tunnus ei ole automaatselt suurema mõjuga kui väikse väärtuste vahemikuga (näiteks  $0 \dots 1$ ) tunnus.

## 4 Andmete analüüs

Töös uuritakse, kui paljud geenidoonorid valimist saavad kümne aasta jooksul ehk aastast 2013 kuni aastani 2023 astma diagnoosi ning kas ja kuidas on astmadiagnoosi saamine seotud väliskeskkonnaga. Analüüsis eeldatakse, et keskkonnatingimused kümne aasta jooksul oluliselt ei muutu ning et valimis olevad isikud elavad aastast 2013 kuni aastani 2023 samas kohas. Analüüsiks sobivas valimis on isikud, kes on vaatlusperioodi alguseks ehk enne 2013. aastat ilma astmadiagnoosita ning elus.

Töös on kasutatud Tartu Ülikooli Eesti Geenivaramu, Eesti Haigekassa Ravikindlustuse andmekogu ja EXPANSE projekti poolt mudeldatud keskkonnaandmeid. Uuring toimus EBINI loa "Mõjusfääri-põhised tööriistad tervisliku linnakeskkonna heaks"(1.1-12/388 (13.02.2024), 1.1-12/3435 (08.12.2020), 1.1-12/1021 (13.04.2021), 1.1-12/1021 (14.12.2021), 1.1-1/3452 (20.10. 2022), 1.1-12/1086 (13.03.2023), 1.1-12/4367 (07.12.2023)) alusel, andmed väljastati andmeväljastuse R24 alusel ning analüüsiti Tartu Ülikooli serveris.

### 4.1 Üldtunnuste valik

Lisas 2 on välja toodud andmestiku üldtunnused, mida on kasutatud käesolevas töös. Iga geenidoonori kohta on olemas info soo, sünniaasta kui ka Geenivaramuga liitumise aja kohta. Sellest võetakse mudelisse tunnused *sugu*, *vanus* ja *liitumisperiod*. Pärast liitumist ja enne 2023. aastat surnud geenidoonorite kohta on olemas surmakuupäev. Nendel geenidoonoritel, kes on astmaatikud (vaata peatükki 1), on olemas vastava diagnoosi RHK-10 klassifikaatori kood kui ka diagnoosi saamise kuupäev. Töös kasutatakse kõigest esimese diagnoosi kuupäeva ehk *esma\_kp*.

Väärtused tunnuste *bmi* (kehamassiindeks), *suits\_staatus* (suitsetamise staatus) kui ka *haridus* kohta on saadud küsimustikest, millele on iga geenidoonor liitumisel

või pärast liitumist vastanud. Osa geenidoonoreid on kutsutud tagasi erinevate uuringute raames, mille tõttu mitmed aastad tagasi liitunud geenidonoritel võib olla kokku kuni neli erinevat väärtust iga tunnuse kohta. Kui geenidonoril on nendes tunnustes mitu väärtust, siis antud töös kasutatakse selle küsimustiku vastust, mis on võimalikult lähedal vaatluse algusele ehk 2013. aastale. Kuigi neil tunnustel on üle 45 000 puuduva väärtuse, siis need sisaldavad olulist infot mudelile, sest kõik kolm on varasemate uuringute ([4], [20] ja [21]) põhjal osutunud astmariski mõjutavateks teguriteks. Antud töös vaadeldakse suitsetamist binaarse tunnusena *suits*, kus praeguseid ja endiseid suitsetajaid ei eristata omavahel. Tunnused *suits*, *haridus* ja *bmi* kaasatakse mudelisse.

## 4.2 Eksposoomitunnuste valik

Analüüsiks on kasutada eksposoomiandmestik, kus ühe geenidoonori kohta esineb kuni nelja erineva elukoha väliskeskkonna andmeid. Analüüsijale pole nähtav info aadresside ega koordinaatide kohta, vaid erinevate geenidoonori aadressidega lingitud keskkonnanäitajate väärtused. Linkimiseks kasutatav elukoht on saadud esmajärjekorras geenidoonori enda öeldud aadressi põhjal või selle puudumisel rahvastikuregistrist.

Ühe geenidoonori potentsiaalne keskkonnaandmete paljusus tuleneb sellest, et Eesti Geenivaramu fikseerib oma süsteemis geenidoonori elukoha iga kord, kui geenidonor küsimustikule vastab. Mitmeid aastaid tagasi liitunud geenidonoritel on üldandmeid küsitud mitmel korral ehk on ka mitmeid keskkonnaandmeid, mis on isikuga seotud. Antud töös kasutan mitme eksposoomiandmete hulga korral andmeid, mille väärtused on lähedal vaatlusaja algusele ehk aastale 2013.

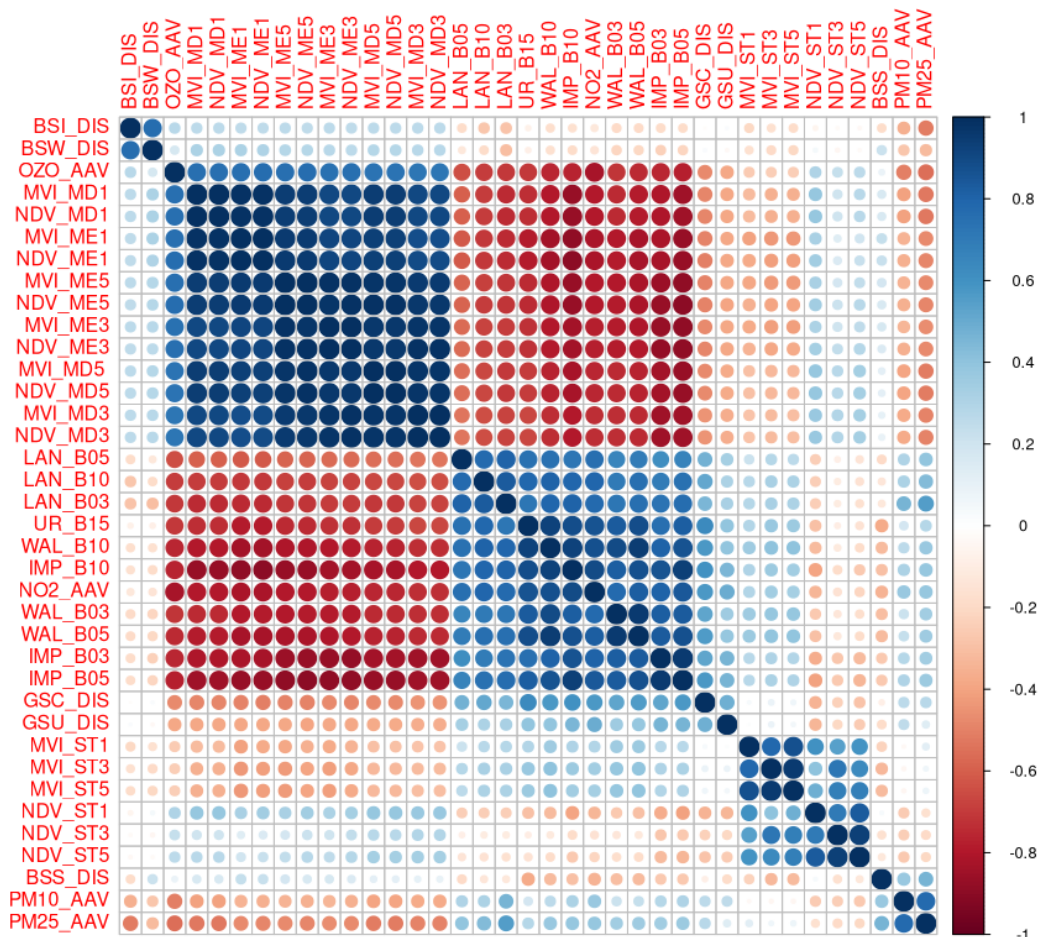
Keskkonnaandmetes on antud 37 erinevat tunnust (vt Lisa 3). Sisuliselt on mõned tunnused väga sarnased, sest nad erinevad teineteisest ainult selle poolest, kui suure raadiuses neid on mõõdetud. Näiteks tunnused *IMP\_B03*, *IMP\_B05* ja *IMP\_B10* näitavad kõik kodu ümbruses olevate kõvade pindade hulka, kus

suuremad väärtused esinevad linnades ning väiksemad väärtused maapiirkondades. Selliste tunnuste samaaegne vaatlemine võib tekitada mudelisse vastuolusid, mille tõttu peaks igast sisuliselt sarnasest tunnuste plokist valima analüüsi ainult ühe.

Iga keskkonnatunnuse kohta on olemas väärtused erinevatest aastatest. Õhureostusega seotud tunnuste, nagu näiteks *NO2\_AAV*, kohta on kakskümmend aastat järjest iga aasta keskmised olemas, kuid taimestikuga seotud tunnustel (*MVI* või *NDV* algusega) on väärtused olemas iga viie aasta tagant, kuna maakasutus ei muutu nii kiiresti. Kõigi tunnuste peale pole ühte kindlat aastat, millal igal keskkonnatunnusel on väärtused olemas. Antud töös valitakse igale keskkonnatunnusele väärtused, mis on võimalikult 2013. aasta läheduses, sest siis on analüüsitav, kuidas fikseeritud keskkonnatingimused vaatluse algul mõjutavad valimis olevate isikute astmasse haigestumist terve ülejäänud vaatluse jooksul.

Lisas 3 on antud ka igal tunnusel esinevate puuduvate väärtuste arv kogu 182 107 geenidoonori seas. Tunnusel *GSU\_DIS* on puuduvaid väärtusi üle 70 000, sest see on defineeritud ainult linnas. *GSU\_DIS* jäetakse mudelist välja, kuna selle kasutamine vähendaks oluliselt valimi suurust ning töös uuritakse ka geenidoonoreid, kes ei ela linnas. Samamoodi on liiga palju puuduvaid väärtusi öise tehisvalguse hulka iseloomustavatel tunnustel (*LAN* algusega), mis jäetakse mudelist välja. Kõikidel teistel tunnustel kõigub puuduvate väärtuste arv 4000 ümber. Sisuliste plokkide sees ei esine suuri erinevusi tunnuste vahel ehk puuduvate väärtuste osas ei ole olulist vahet, kas mudelisse on valitud näiteks tunnus *IMP\_B03*, *IMP\_B05* või *IMP\_B10*. Kõigi selliste tunnuste puhul valitakse töös mudelisse tunnus, mis on mõõdetud kõige suurema ehk 1000 m raadiuse maa-ala kohta. Valik on tingitud sellest, et valimis olevad isikud on kõik täiskasvanud ja arvatavasti liiguvad suuremal raadiusel oma kodu ümber ringi kui kõigest 300 või 500 meetrit ning kogu 1000 meetri raadius ilmestab paremini nende igapäevast eksposoomi kokkupuudet. Selleks, et mudelis ei tekiks multikollineaarsust, on vaja arvestada tunnustevaheliste seostega. Kuna mitmed keskkonnatunnused ei ole normaaljaotusega (vt Lisa 4),

siis on sobilik tunnustevahelisi seoseid mõõta Spearmani astakorrelatsioonikordajaga. Seosekordajaid näitaval joonisel 1 on näha, et esinevad tunnuste plokid, kus omavaheline korrelatsioon on tugev või lausa väga tugev.



Joonis 1: Eksposoomitunnuste korrelatsioonimaatriks.

Joonis näitab tugevat seost tunnuste *BSI\_DIS* ja *BSW\_DIS* vahel, mis tähendab et lähima siseveekogu kaugus kirjeldab ka lähima (sise- või välis-)veekogu kaugust eluasemest. Kuid nendel kahel tunnusel pole tugevat seost tunnusega *BSS\_DIS* ehk merepiiri kaugusega. Mudelisse valitakse veekogude kaugusi elukohast ilmes-tama kaks teineteisest erinevat tunnust *BSI\_DIS* ja *BSS\_DIS*. Välja jäetakse tunnus *BSW\_DIS* selleks, et ei tekiks vastuolusid tunnusega *BSI\_DIS*.

Selgelt on näha joonisest, et noore taimestiku osakaalu aastase mediaani kui ka keskmise (*MIV\_MD* ja *MVI\_ME* algusega tunnuste) seosekordaja täiskasvanud taimestiku osakaalu aastase mediaani kui ka keskmise (*NDV\_MD* ja *NDV\_ME* algusega tunnuste) vahel on ühe lähedane, mille tõttu aastast taimestiku osakaalu on mõistlik ilmentada neist kõigist vaid ühe tunnusega. Sobivaks tunnuseks on valitud *NDV\_ME1*, sest täiskasvanud taimestik tekitab õitsemisel õhku õietolmu osakesi. Aastase mediaani asemel vaadeldakse aastast keskmist, sest see sisaldab rohkem infot ekstreemumitest. Sellele lisaks on mudelisse kaasatud antud tunnuse aastane standardhälve ehk *NDV\_ST1*.

Joonis ilmestab, et ka mitmetel teistel tunnustel, mis on sisuliselt teineteisest erinevad, on tugevaid positiivseid kui ka negatiivseid omavahelisi seoseid. Töös arvestatakse antud seostega, kuid rohkem tunnuseid selle tõttu välja ei jäeta.

Mudelisse valitud tunnused ja aastad on märgitud lisas 3 kollase värviga. Valimisel on arvestatud kõikide selle peatüki eelnevate aruteludega.

### 4.3 Valim

Analüüsiks on kasutada 182 107 Tartu Ülikooli Eesti geenivaramu kogutud geenidoonori andmed, kellest 12 949 on astmaatikud (vaata peatükki 1). Valimi moodustamiseks filtreeritakse välja 7791 geenidoonorit, kellel on astma esmadiagnoos antud enne 2013. aastat. Samuti ei kaasata valimisse 1506 astmata geenidoonorit, kes on surnud enne 2013. aastat. Analüüsis ei saanud ka kasutada 4500 doonori andmeid, kelle elukoha kohta valitud eksposoomitunnuste andmed puuduvad. Viimaks on eemaldatud valimist ka 50 384 geenidoonorit, kes pole vastanud küsimustikule ehk puuduvad andmed hariduse, suitsetamise või kehamassiindeksi kohta.

Lõplikus valimis on 118 594 geenidoonorit ning nendest saavad kümne aasta jooksul astma esmadiagnoosi 3574 isikut.

Kõikide valimis olevate isikute kohta on teada 8 üldist ja 12 eksposoomi kirjeldavat



tunnust. Kõik need tunnused on kasutusel järgnevas Coxi mudelis. (Vaata Lisa 2 ja Lisa 3 kollaseid tunnuseid.)

## 4.4 Coxi mudel

Kasutatav mudel peab arvestama paremalt tsenseeritusega, sest valimis leidub geenidoonoreid, kellel vaatlusaja lõpuks ei diagnoositud astmat, kuid kes võivad saada astma diagnoosi hiljem. Analüüsimiseks on sobiv Coxi võrdeliste riskide mudel. Mudelis võrreldakse teatud ajahetkel astma diagnoosi saanute väliskeskkonna andmeid nende geenidonorite väliskeskkonna andmetega, kes selleks hetkeks ei ole saanud veel astma esmadiagnoosi. Sellisel võrdlusel saab leida, millised mudelisse pandud tunnused suurendavad või vähendavad astmariski statistiliselt olulisel määral. Mudelis uuritakse aega kuni astma esmadiagnoosini. Vastav sündmus on fikseeritud uue tunnusega nimega *astma* (vt Lisa 2), mille väärtus 0 tähistab tsenseeritud ja väärtus 1 tsenseerimata vaatluseid. Sündmuse toimumiseni kulunud aeg on fikseeritud tunnusega *aeg* (vt Lisa 2), mis on tunnuse *esma\_kp* väärtuse ja vaatluse algusaja ehk 01.01.2013 vahe aastates. Nende puhul, kes ei saanud vaatlusaja jooksul astmadiagnoosi ehk sündmus ei toimunud, on sündmuse toimumise aeg tsenseeritud. Tsenseeritud vaatluste puhul on *astma* väärtus 0 ning *aeg* on 10 aastat või *surma* kuupäeva olemasolul selle ja 01.01.2013 vahe aastates. Mudelis kasutatav olulisuse nivoo on 0,05. Pidevad tunnused on mudelis skaleeritud ehk pideva tunnuse ühikuks on üks standardhälve.

Esmalt lisatakse mudelisse valitud kuus üldtunnust ühekaupa (1. mudeli kuju on  $(aeg, astma) \sim tunnus$ ) ning fikseeritakse iga tunnuse p-väärtus (vt tabel 1). Tabelites tarkvara R funktsioon *scale()* tähistab pideva tunnuse skaleerimist.

Esimese mudeli väikseima p-väärtusega tunnus on *bmi*, mis jäetakse järgmisse mudelisse alles ehk 2. mudeli kuju on  $(aeg, astma) \sim scale(bmi) + tunnus$ . Igas järgnevas mudelis on eelmiste mudelite vähima p-väärtusega tunnused, millele lisatakse ühekaupa alles jäänud tunnused. Tunnuseid lisatakse mudelisse kuni kõigi

Tabel 1: Üldtunnuste p-väärtused mudelites.

Tunnuse lühend	1. mudeli p-väärtus	2. mudeli p-väärtus	3. mudeli p-väärtus	4. mudeli p-väärtus
<i>sugu</i>	1,24e-09	1,51e-13		
<i>scale(vanus)</i>	6,01e-05	0,923	0,547	0,662
<i>liitumisperiod</i>	0,0784	0,141	0,136	0,954
<i>scale(bmi)</i>	<2e-16			
<i>suits</i>	0,105	0,457	0,0294	0,151
<i>haridus</i>	6,36e-11	1,45e-08	5,47e-10	

uute üksikute tunnuste p-väärtused on üle olulisuse nivoo.

Tunnused *sugu*, *bmi* ja *haridus* osutuvad üldtunnuste mudelis statistiliselt olulisteks. Tarkvara R funktsiooniga *cox.zph* tehakse saadud üldtunnuste mudelile võrdeliste riskide eelduste kontroll, millest ilmneb, et tunnus *haridus* ei vasta eeldustele. Muutes tunnuse *haridus* mudelis järjestustunnuse asemel binaarseks tunnuseks, kus haridus jaotub kaheks: kõrgharitud ning ilma kõrghariduseta isikud, saadakse mudelisse uuesti statistiliselt oluline tunnus, mis täidab ka kontrollis mudeli eeldused.

Järgnevalt lisatakse valmis üldtunnuste mudelisse ühekaupa 12 erinevat eksposoom-tunnust (vt tabel 2). Nüüd on 1. mudeli kujuks  $(aeg, astma) \sim scale(bmi) + sugu + I(haridus > 6)TRUE + tunnus$ . Tunnuste valikus järgnevatesse mudelitesse järgitakse sama strateegiat, mis üldtunnuste mudeli koostamisel.

Tunnused *BSS\_DIS*, *GSC\_DIS* ja *OZO\_AAV* osutuvad eksposoomitunnuste mudelis statistiliselt olulisteks. Tarkvara R funktsiooniga *cox.zph* tehakse saadud mudelile võrdeliste riskide eelduste kontroll, millest ilmneb, et tunnus *BSS\_DIS* ei vasta eeldustele. Muutes tunnuse *BSS\_DIS* mudelis pideva tunnuse asemel binaarseks tunnuseks, kus elukoha kaugus merest jagatakse mediaani põhjal kaheks väärtuseks: elukoht on merest alla mediaani (10,9 km) kaugusel ja elukoht on merest üle mediaani kaugusel, saadakse mudelisse uuesti statistiliselt oluline tunnus, mis täidab ka kontrollis mudeli eeldused.

Tabel 2: Eksposoomitunnuste p-väärtused mudelites.

Tunnuse lühend	1. mudeli p-väärtus	2. mudeli p-väärtus	3. mudeli p-väärtus	4. mudeli p-väärtus
<i>scale(BSI_DIS)</i>	0,00145	0,0969	0,017763	0,147202
<i>scale(BSS_DIS)</i>	1,93e-11			
<i>scale(GSC_DIS)</i>	5,69e-05	0,003148		
<i>scale(IMP_B10)</i>	0,306	0,723	0,00472	0,84893
<i>scale(NDV_ME1)</i>	0,333	0,964	0,035359	0,654449
<i>scale(NDV_ST1)</i>	0,631	0,709551	0,335619	0,953
<i>scale(NO2_AAV)</i>	0,133	0,704	0,00664	0,222665
<i>scale(OZO_AAV)</i>	0,0551	0,0309	5,46e-05	
<i>scale(PM10_AAV)</i>	0,00216	0,722	0,028399	0,62923
<i>scale(PM25_AAV)</i>	3,74e-06	0,491	0,008835	0,30637
<i>scale(UR_B15)</i>	0,197	0,553	0,00119	0,27583
<i>scale(WAL_B10)</i>	0,91	0,178	0,00146	0,34026

Tabelis 3 on näha lõplikud tulemused, kus tarkvara R funktsioon  $I()$  tähistab järjestus- või pideva tunnuse binaarseks tunnuseks muutmist.

Tabel 3: Astmariski mõjutavad tunnused.

Tunnuse lühend	$e^{\hat{\beta}}$	95% usaldusintervall	p-väärtus
<i>scale(bmi)</i>	1,23	(1,20; 1,27)	<2e-16
<i>sugu</i>	1,33	(1,24; 1,43)	2,41e-14
<i>I(haridus &gt; 6)TRUE</i>	0,87	(0,81; 0,93)	0,000103
<i>I(BSS_DIS &gt; 10900)TRUE</i>	1,25	(1,17; 1,34)	2,43e-10
<i>scale(GSC_DIS)</i>	0,92	(0,88; 0,96)	7,34e-05
<i>scale(OZO_AAV)</i>	0,91	(0,88; 0,95)	2,30e-06

Astmariski suurendab kehamassiindeksi väärtuse suurenemine. Kuna valimi kehamassiindeksi standardhälve on 5 ühikut, siis näiteks isikul, kelle kehamassiindeks on 25, on 1,23 korda suurem tõenäosus saada astma kui isikul, kelle kehamassiindeks on ühe standardhälbe võrra madalam ehk 20. Riski suurendab ka naissugu. Mudeli hinnangu põhjal on naistel 1,33 korda suurem tõenäosus saada astma kui meestel. Eksposoomitunnustest suurendab astma riski elukoha kaugus merest. Pooled valimi liikmed elavad merest kaugemal kui 10,9 kilomeetrit ning mudeli hinnangu põhjal

neil on 1,25 korda suurem risk astmasse haigestuda kui teisel poolel valimist, kes elavad merele lähemal kui 10,9 kilomeetrit. Analüüsi tulemuste põhjal vähendab astma riski bakalaureuse-, magistri- või doktorikraadi omamine. Kõrgeltharitud isikul on  $\frac{1}{0,87} = 1,15$  korda väiksem tõenäosus astmasse haigestuda kui madalama haridusega isikul. Eksposoomitunnustest vähendab astma riski kaugus CORINE'i rohealadest. Kuna valimi CORINE'i rohealade kauguse standardhälve on 336 meetrit, siis isikul, kes elab rohealast 336 meetrit kaugemal, on  $\frac{1}{0,92} = 1,09$  korda väiksem tõenäosus astmasse haigestuda kui teisel isikul, kes elab lähemal. Näiteks, kui elukoht on lähimast rohealast 1 kilomeetri kaugusel, siis see vähendab astma riski 3,27 korda võrreldes elukohaga, mis on otse roheala kõrval ehk kaugus on alla kümne meetri. Interpreteerides CORINE'i rohealade kauguse mõju astmariskile, peab arvestama, et CORINE'i rohealade alla kuuluvad ainult suure pindalaga alad ning et sinna kuuluvad väga erineva karakteristikutega piirkonnad (vaata peatükki 2.2). Astma riski vähendab suurem aastane keskmine osooni hulk õhus. Osooni aastase keskmise hulga standardhälve antud valimis on 3 mikrogrammi kuupmeetri kohta. Seega 3-mikrogrammine tõus osooni aastases keskmises hulgas vähendab astma riski  $\frac{1}{0,91} = 1,10$  korda. Näiteks valimi madalaim aastane keskmine osooni hulk õhus on ligikaudu 33 mikrogrammi kuupmeetri kohta, kuid kõrgeim on 69 mikrogrammi kuupmeetri kohta. Osooni maksimaalse väärtusega elukoha elanikul on saadud mudeli põhjal ligikaudu 13 korda väiksem astma risk kui minimaalse väärtusega elanikul. Valimil on osooni kõrged väärtused maapiirkonnas ning madalamad väärtused linnas, sest tunnus *OZO\_AAV* on positiivselt korreleeritud rohelist näitavate keskkonnatunnustega, nagu *NDV\_ME1*, ning negatiivselt korreleeritud linnakeskkonda ilmestavate tunnustega, nagu *UR\_B15*, *NO2\_AAV*, *IMP\_B10* (vaata peatüki 4.2 joonist 1).

## 4.5 Peakomponentanalüüs

Eelnevas peatükis 4.4 on tulemustes eraldi iga tunnuse tõlgendamine keeruline, sest peab ka arvestama, et eksposoomitunnused on omavahel tihedalt seotud (vaata peatükis 4.2 joonist 1). Tulemusi interpreteerides on õigem vaadelda keskkonnatunnuste komplektide mõju astmariskile. Keskkonnatunnuste omavaheliste seoste paremaks arvestamiseks on mõistlik kasutada peakomponentanalüüsi, kus leitavate peakomponentidega saab samuti moodustada Coxi mudeli.

Peakomponentanalüüs on läbi viidud kaheteistkümne valitud eksposoomitunnuse standardiseeritud väärtustega, kus keskvärtus on null ja standardhälve üks. Tabelist 4 on näha, et vähemalt 95% valitud eksposoomitunnuste väärtuste hajuvusest kirjeldab ära kaheksa esimest peakomponenti kaheteistkümnest. Esimesed neli peakomponenti kirjeldavad vähemalt 80% hajuvusest.

Tabel 4: Esimese kaheksa peakomponendi varieeruvuse kirjeldamine.

	<b>PC1</b>	<b>PC2</b>	<b>PC3</b>	<b>PC4</b>	<b>PC5</b>	<b>PC6</b>	<b>PC7</b>	<b>PC8</b>
standardhälve	2,503	1,519	0,883	0,835	0,775	0,663	0,547	0,471
osakaal koguvarieeruvusest	0,522	0,192	0,065	0,058	0,050	0,037	0,025	0,018
kumulatiivne varieeruvus	0,522	0,715	0,780	0,838	0,888	0,924	0,949	0,968

Järgnevalt on esmalt hinnatud Coxi mudel, kus on argumentideks vaid üldtunnused ja seejärel on sinna ükshaaval lisatud peakomponente:  $(aeg, astma) \sim scale(bmi) + sugu + I(haridus > 6)TRUE + peakomponent$ . Teise mudeli edasiarendamisel järgitakse tunnuste valikul sama strateegiat, mis on kirjeldatud peatükis 4.4. Mudelisse saab mitut peakomponenti korraga lisada, sest nende väärtused pole omavahel korreleeritud. Tabelist 5 on näha, kuidas lõpliku mudelini on jõutud.

Peakomponendid kaks, kolm ja viis osutuvad mudelis statistiliselt olulisteks. Tehes saadud mudelile võrdeliste riskide eelduste kontroll ilmneb, et peakomponent kaks ei vasta mudeli eeldustele. Muutes teise peakomponendi mudelis pideva tunnuse

Tabel 5: Peakomponentide p-väärtused mudelites.

Lisatud pea- komponent	1. mudeli p-väärtus	2. mudeli p-väärtus	3. mudeli p-väärtus	4. mudeli p-väärtus
<i>PC1</i>	0,796	0,660928	0,688	0,70428
<b><i>PC2</i></b>	<b>4,14e-10</b>			
<b><i>PC3</i></b>	0,00464	0,00344	<b>0,00284</b>	
<i>PC4</i>	0,253	0,235	0,221	0,2461
<b><i>PC5</i></b>	6,30e-06	<b>8,25e-06</b>		
<i>PC6</i>	0,304	0,344	0,413	0,40617
<i>PC7</i>	0,0574	0,0589	0,086	0,08337
<i>PC8</i>	0,341	0,366	0,373	0,35808

asemel binaarseks tunnuseks, kus väärtused on jagatud keskmise põhjal kaheks: teise peakomponendi väärtus on alla keskmise (0) ja väärtus on üle keskmise, saadakse mudelisse uuesti statistiliselt oluline tunnus, mis täidab ka kontrollis mudeli eeldused. Tabelis 6 on näha lõpliku mudeli tulemused.

Tabel 6: Astmariski mõjutavad tunnused peakomponente kaasavas mudelis.

Tunnuse lühend	$e^{\hat{\beta}}$	95% usaldusintervall	p-väärtus
<i>scale(bmi)</i>	1,24	(1,20; 1,27)	<2e-16
<i>sugu</i>	1,33	(1,24; 1,43)	1,70e-14
<i>I(haridus &gt; 6)TRUE</i>	0,86	(0,80; 0,92)	3,43e-05
<i>I(PC2 &gt; 0)TRUE</i>	0,82	(0,91; 0,95)	9,79e-09
<i>PC5</i>	1,11	(1,06; 1,16)	1,56e-06
<i>PC3</i>	0,94	(0,91; 0,98)	0,00351

Tulemuste tõlgendamiseks on tarvilik vaadelda, millest peakomponendid koosnevad. Selleks on tabelis 7 välja toodud iga 12 eksposoomitunnuse kordajate suurused, mis näitavad kui suure ja millise suunaga mõju on antud eksposoomitunnusel kindlale peakomponendile. Tabelis 7 on kollasega märgitud keskkonnatunnuste kordajad, mis mõjutavad kõige rohkem peakomponente ehk mille kordaja väärtus on vähemalt 0,40. Analüüsidest peakomponentide tunnuste kordajaid koos saab leida otsitavad keskkonnatunnuste komplektid ja nende mõju astmale.

Tabel 7: Eksposoomitunnuste kordajad peakomponentidel.

Tunnuse lühend	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
<i>BSI_DIS</i>	0,15	0,31	0,54	0,76	0,08	-0,03	0,01	-0,01
<i>BSS_DIS</i>	0,05	-0,57	0,03	0,23	0,20	0,53	-0,32	0,16
<i>GSC_DIS</i>	-0,28	0,05	0,35	-0,10	-0,78	0,18	-0,23	0,16
<i>IMP_B10</i>	-0,38	0,11	-0,04	0,08	0,07	0,15	0,19	0,02
<i>NDV_ME1</i>	0,35	-0,05	0,11	-0,13	-0,03	-0,08	-0,50	-0,65
<i>NDV_ST1</i>	0,23	0,03	-0,69	0,49	-0,46	-0,04	-0,00	-0,04
<i>NO2_AAV</i>	-0,37	0,11	-0,06	0,07	0,05	-0,29	-0,10	-0,07
<i>OZO_AAV</i>	0,34	0,03	0,16	-0,16	-0,23	0,40	0,61	-0,29
<i>PM10_AAV</i>	-0,23	-0,45	0,17	0,12	-0,13	-0,46	0,24	-0,38
<i>PM25_AAV</i>	-0,19	-0,53	0,04	0,19	-0,10	0,03	0,18	-0,06
<i>UR_B15</i>	-0,35	0,19	-0,09	0,02	-0,07	0,30	-0,25	-0,35
<i>WAL_B10</i>	-0,34	0,16	-0,16	0,13	0,22	0,34	0,14	-0,41

Esimest peakomponenti mõjutavad mõõdukalt mitmed keskkonnatunnused: kõvade pindade hulk, roheluse indeks NDV, lämmastikdioksiidi ja osooni kogus õhus, linnakvaliteedi ja jalutamiseks sobivuse indeksid. Esimene peakomponent kirjeldab kordajate põhjal elukeskkonna erinevust maal ja linnas. Coxi mudeli põhjal pole maal või linnas elamisel olulist erinevust astmariski mõjule.

Teise peakomponendi väärtust mõjutavad kordajate põhjal kõige rohkem kaugus merest, peenete ja ülipeenete osakeste hulk õhus (tabelis 7 kollasega märgitud kordajate väärtused). Mudeli põhjal teisel peakomponendil on oluline mõju astmariskile. Teine peakomponent kirjeldab, kuidas astmariski vähendab elamine elukohas, kus meri on lähedal ning peenosakesi (nii  $PM_{10}$  kui ka  $PM_{2,5}$ ) on õhus vähe. Saab ka öelda, et astmariski suurendab elukoht, mis on merest kaugel ja kus on õhus palju peenosakesi.

Kolmandat peakomponenti mõjutavad kõige rohkem roheluse indeksi NDV aastane standardhälve ja kaugus siseveekogust. Kolmanda peakomponendi põhjal vähendab astmariski elukoht, kus siseveekogu on kaugel ning rohelusel on väike aastane standardhälve ehk näiteks okaspuude lähedal elamine. Vastukaaluks saab ka öelda, et

astmariski suurendab elukoht, mis on siseveekogu lähedal ning kus rohelusel on suur aastane standardhälve ehk näiteks lehtpuud või põllud on lähedal.

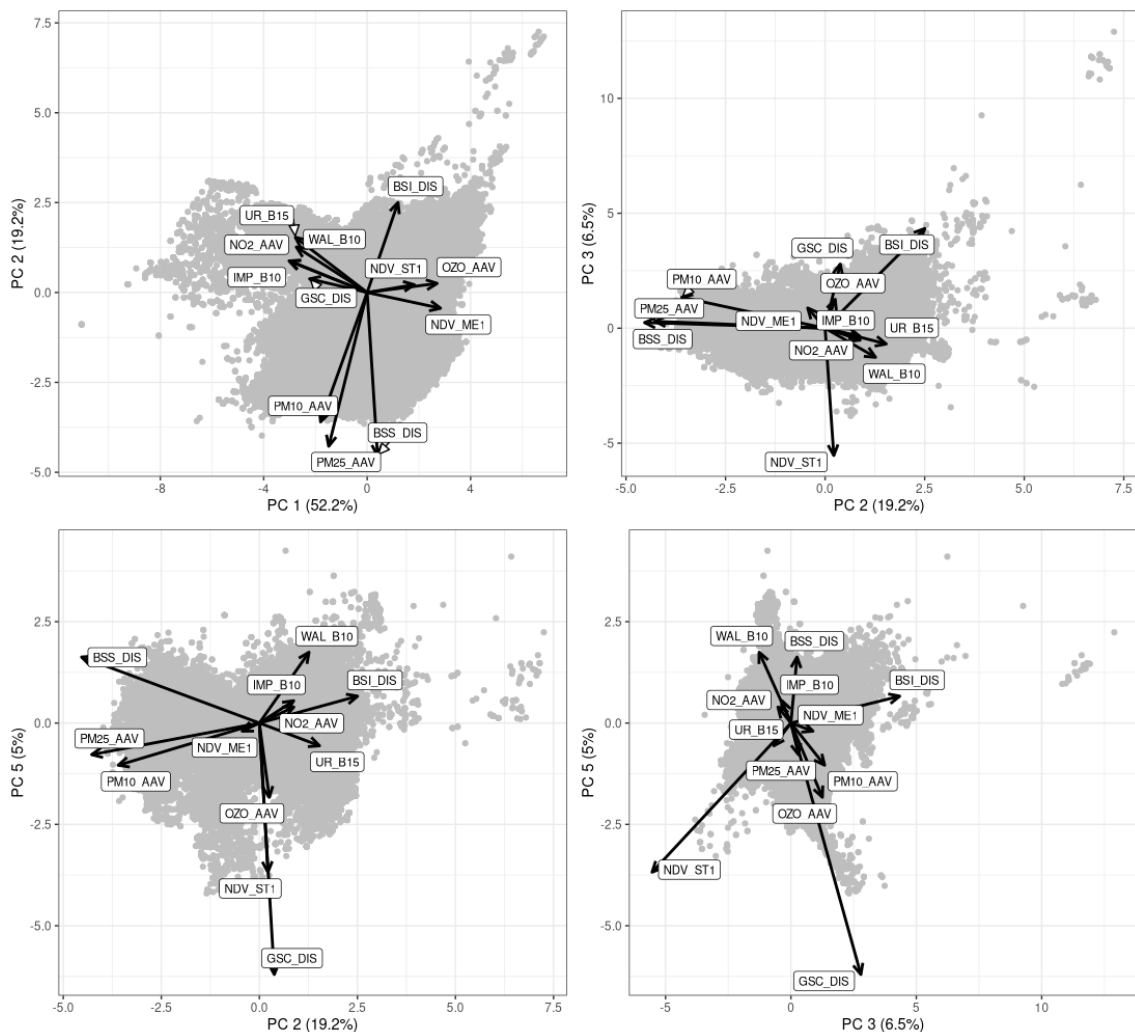
Neljandat peakomponenti mõjutavad samuti kõige rohkem roheluse indeksi NDV aastane standardhälve ja kaugus siseveekogust, kuid see peakomponent ei ole Coxi mudelis oluline. See tähendab, et elukeskkonnal, kus siseveekogu on kaugel ning aastane roheluse indeksi standardhälve on suur ehk näiteks lehtpuud või põllud on lähedal, pole olulist mõju riskile saada astmadiagnoos. Astmariski ei mõjuta samuti elukeskkond, kus siseveekogu ja okaspuud on lähedal.

Viiendat peakomponenti mõjutavad oluliselt CORINE'i rohealade kaugus ja roheluse indeksi NDV aastane standardhälve. Viies peakomponent on mudelis oluline ning selle põhjal astmariski suurendab elukoht, mis on lähedal CORINE'i rohealadele (vaata peatükki 2.2) ja kus on rohelusel väike aastane standardhälve ehk kus okaspuud on lähedal. Lisaks saab öelda, et astmariski vähendab elukeskkond, kus CORINE'i rohealad on kaugel ning lehtpuud või põllud on lähedal.

Kuues, seitsmes ega ka kaheksas peakomponent ei ole olulised Coxi mudelis ehk nende kirjeldatud keskkondadel ei ole olulist mõju astmariskile.

Mudelis olulisteks osutunud peakomponentide kaks, kolm ja viis kordajate väärtusi ilmestab graafiliselt ka joonis 2. Joonisel on samuti olemas graafik esimese ja teise peakomponendi väärtuste kohta, sest mõlemad kokku kirjeldavad valimi eksposoomitunnuste väärtuste hajuvusest üle 70%.





Joonis 2: Esimese, teise, kolmanda ja viienda peakomponendi väärtused.

Tabelist 7 leitavate kordajate ja joonise 2 põhjal saab soovi korral interpreteerida veelgi täpsemalt igat peakomponenti, mitte vaadelda ainult tunnuseid, mille kordajate väärtused on üle 0,40. Võrreldes eelnevalt leitud Coxi mudeliga peatükis 4.4 on siinse peakomponentanalüüsi tulemused sarnased kauguse merest ja CORINE'i rohealadest osas. Peakomponentanalüüs on andnud tulemustele juurde siseveekogude kauguse, peenete- ja ülipeenete osakeste hulga õhus, roheluse indeksi NDV aastase standardhälve mõju astmariskile. Peakomponentanalüüs on näidanud, et kogu väliskeskkonda on vaja analüüsida tervikuna, sest sama keskkonnatunnus võib

erinevates kombinatsioonides olla erineva mõjuga astmariskile. Näiteks tulemustest tuleb välja, et roheluse väike aastane standardhälve ehk okaspuude lähedal elamine mõjutab astmariski erinevalt erinevates keskkondades: siseveekogust kaugel olles vähendab riski, CORINE'i rohealade kõrval elades suurendab riski.

## 5 Tulemuste arutelu

Tulemustest selgub, et töös kasutataval valimil naissugu suurendab astmariski, mis läheb kokku varasemate leidudega, et naissuguhormoonid suurendavad hingamisteede põletiku teket ning testosteroon vähendab [22]. Valimil kõrge kehamassiindeks suurendab astmariski, mis on ka oodatav tulemus, sest varasemalt on leitud, et näiteks rasvunud lastel on suurem risk astmadiagnoos saada ning neil on ka raskemad astma sümptomid [21]. Mudeli põhjal kõrgharidus vähendab astmariski. Samasuguse tulemuse sai ka Eagan 2004. aastal ning annab sellele potentsiaalseks põhjenduseks, et madalama haridusega täiskasvanutel võivad olla näiteks halvemad toitumisharjumused ning eluruumide tingimused [20]. Suitsetamise riskifaktori mõju siinses töös ei õnnestunud näidata, kuigi see on näiteks Levy 2015. aasta uuringu põhjal üks suurimaid hingamisteede ärritajaid [4]. Samuti ei osutunud oluliseks riskifaktoriks astmale vanus, kuigi leidub uuringuid, kus on leitud, et astma risk suureneb täiskasvanutel vanuse kasvades [23].

Kuna eksposoomitunnuste korrelatsioonimaatriks joonisel 1 näitas, et mitmed valitud eksposoomitunnused on tihedalt omavahel korreleeritud, siis tulemuste interpreteerides on õigem vaadelda keskkonnatunnuste komplektide mõju astmariskile ehk tõlgendada peakomponentanalüüsi tulemusi.

Peakomponentanalüüsi tulemuste põhjal astmariski osas pole vahet, kas vaatlusalune elab maal või linnas. Malik 2012. aastal erinevaid astma ja keskkonnaga seotud uuringuid retrospektiivselt analüüsid andis selle nähtuse põhjenduseks, et mitmetel maa- ja linnapiirkondadel on arvatavasti sarnased keskkonna- ja sotsiaalmajanduslikud probleemid, mis mõjutavad astmat [24]. Kuigi mitmed uuringud on tuvastanud maa ja linna piirkondadel erinevuse astmariskis ning ka Maailma Terviseorganisatsioon seostab astmariski linnastumise ja sealse õhusaastega [6], siis antud valimi näitel Eesti elanikkonnas pole olulist vahet astmariskil linnas ja maal elanute seas.

Astmariskile on oluline vähendav mõju elukeskkonnal mere ääres, kus on õhus vähe

peenosakesi (nii  $PM_{10}$  kui ka  $PM_{2,5}$ ). Varasemate uuringute põhjal on oodatav, et astmariski vähendav mõju on peenosakeste vähesel kogusel õhus [12]. Saudi-Araabias on ka leitud Beneri poolt 1993. aastal, et astma sümptomeid on rohkem sisemaal elanud lastel kui mere ääres elanutel [25]. Mere kauguse mõju astmariskile võib põhjendada sellega, et seal hajub õhureostus lihtsasti merele ning see tõttu on astmaatikutele õhus vähem ärritajaid.

Tulemuste põhjal astmariski vähendab oluliselt elukeskkond, kus siseveekogu on kaugel ning okaspuud on lähedal. Põhjus võib olla selles, et enamasti on siseveekogude ümber org, kust on õhureostusel raskem ära hajuda ning lisaks lehtpuude olemasolul on jõe orus rohkem allergeenset õietolmu, mis ärritab astmaatikuid. Eestis on lehtpuudest allergeense õietolmuga Maailma Terviseorganisatsiooni põhjal harilik sirel, harilik saar, arukask, harilik sarapuu ja sanglepp, kuid allergeensetest okaspuudest leidub Eestis vaid ebaküpressi ilutaimena [26]. Okaspuudest on veel kadakas nõrgalt allergeenne, mida esineb sagedaselt Loode-Eestis ja saartel [27]. Eestis leiduvate teiste okaspuude õietolm ei ole allergeenne ehk enamasti keskkond okaspuude ja siseveekogu ümber pole hingamisteedele küllalt ärritav, et astmariski suurendada, kuid keskkond lehtpuude ja siseveekogude ümber on. 2023. aastal Montreali Kanada lapsi uurinud Duquesne leidis, et õitsemise ajal lehtpuud suurendavad lastel astmariski ning pärast õitsemist suurte lehtede ajal vähendavad astmariski [28].

Keskkond, kus CORINE'i rohealad kui ka okaspuud on lähedal, suurendab töös uuritava valimi põhjal astmariski. Sarnase tulemuse on saanud 2020. aastal ka Parmes Euroopa lapsi uurides [29], ta põhjendab oma vastuolulist leidu rohealade osas sellega, et rohealade alla kuuluvad mitmed eriilmelised maakatte tüübid. Antud töös kuulub CORINE'i rohealade alla samuti erinevate karakteristikutega alad (vaata peatükki 2.2), mille tõttu on raske põhjendada, miks okaspuudega CORINE'i roheala lähedal elamine soodustab astma teket antud valimil. Meeles peab ka pidama, et CORINE'i rohealade täpsus on vaid 25 hektarit. Tartu näitel (vaata lisa 5) on paljude CORINE'i rohealade lähedal samuti ahiküttepiirkonnad, mil-

le tõttu võiks tulevikus lisaks olemasolevatele tunnustele kasutada ka piirkondade küttesüüvide infot, näiteks ahikütte osakaalu.

## Kokkuvõte

Bakalaureusetöö eesmärgiks on leida astma väliskeskkonna riskitegurid ning analüüsida kogu keskkonnategurite komplekti tervikuna. Astmaatikutena on käsitletud inimesi, kellel on vähemalt kaks diagnoosi RHK-10 koodiga J45 või J46 või üks diagnoos koodiga J45 või J46 ning eelnevate koodidega seotud ravimiresept. Analüüsis on arvestatud varasemalt teada olevate astma riskifaktoritega, nagu kehamassiindeks, vanus, sugu, haridus ja suitsetamine. Töös on moodustatud 182 107 Tartu ülikooli Eesti geenivaramu geenidonorist, kellest 12 949 on astmaatikud, valim suurusega 118 594, kes kõik olid ilma astma diagnoosita 2013. aasta alguses. Analüüsis on eeldatud, et valimis olevad inimesed elavad 2013.-2022. aastal ühes kohas ning analüüsitakse, kuidas elukoha väliskeskkond on seotud astma diagnoosi saamisega vaadeldud 10 aasta jooksul. Vaatlusaja jooksul valimist saab 3574 geenidonorit astma esmadiagnoosi.

Coxi mudelis on võrreldud teatud ajahetkel astma diagnoosi saanute väliskeskkonna andmeid nende geenidonorite väliskeskkonna andmetega, kes selleks hetkeks ei ole saanud veel astma esmadiagnoosi. Mudelis on ajaskaalaks astma esmadiagnoosi saamiseks kuluv aeg vaatlusaja algusest 01.01.2013 kuni vaatlusaja lõpuni 31.12.2022. Mudel arvestab paremalt poolt tsenseeritud andmetega.

Tulemustest selgus, et töös kasutataval valimil naissugu ja kõrge kehamassiindeks suurendavad astmariski ning kõrgharidus vähendab astmariski. Vanuse ja suitsetamise riskifaktorite mõju siinses töös ei õnnestunud näidata. Keskkonnatunnused osutusid teineteisega tihedalt korreleerituks ehk on leitud, et neid peaks kompleksitena koos vaatlema, mitte üksikute tunnustena. Selleks kasutati analüüsis peakomponente. Kaheteistkümnest peakomponendist kaheksa kirjeldavad eksposoomitunnuste hajuvust vähemalt 95%. Coxi mudelis osutusid statistiliselt oluliseks peakomponendid kaks, kolm ja viis. Esimesel peakomponendil ei olnud mudelis mõju astmariskile ehk antud valimi põhjal pole vahet, kas vaatlusalune elab maal või linnas. Teise peakomponendi põhjal on astmariskile vähendav mõju elukeskkon-

nal, kus meri on lähedal ning peenosakesi (nii  $PM_{10}$  kui ka  $PM_{2,5}$ ) on õhus vähe. Kolmanda peakomponendi põhjal vähendab siseveekogust kaugel ja okaspuude lähedal elamine valimil astmariski. Viiendast peakomponendist selgus, et CORINE'i rohealade ja okaspuude lähedus suurendab astmariski, kuid kuna CORINE'i rohealad on Tartu linna näitel väga eriilmelised, siis on raske interpreteerida miks okaspuudega CORINE'i rohealad potentsiaalselt tekitavad astmat.

Tulevikus oleks vajalik lisaks peakomponentanalüüsile analüüsida ka eksposüüpe, kus elukeskkond on kirjeldatud erinevate eksposoomikomponentide põhjal, sest kõik eksposoomitunnused mõjutavad inimest korraga. Järgnevates uuringutes oleks soovitatav moodustada klastrid eksposoomi põhjal, näiteks õhureostusega linnapiirkond, puhta õhuga linnapiirkond, õhureostusega roheline piirkond, puhta õhuga roheline piirkond. Klasterdamisel saaks ka näiteks arvestada mere või siseveekogu kaugustega. Olulisi erinevusi võib ka ilmnedu põldude, leht- või okasmetsade lähedal olevates elukeskkondades. Kasulik oleks analüüsi ka lisada uusi tunnuseid, nagu ahi- ja keskküttepiirkonnad või ümbruskonna müratasemed.

## Kasutatud allikad

- [1] Jelle Vlaanderen *et al.* „Developing the building blocks to elucidate the impact of the urban exposome on cardiometabolic-pulmonary disease: The EU EXPANSE project“. *Environ Epidemiol* 5.4 (2021), lk. e162. URL: <https://pubmed.ncbi.nlm.nih.gov/34414346/> (vaadatud 07.04.2024).
- [2] Sotsiaalministeerium. *Rahvusvahelise Haiguste Klassifikatsiooni 10. versiooni klassifikaatorite andmebaas*. 1992. URL: <https://rhk.sm.ee/> (vaadatud 07.04.2024).
- [3] Eesti Tervisekassa. *Kliinilise auditi „Astma käsitus esmatasandil“ kokkuvõte*. 2017. URL: <https://www.tervisekassa.ee/sites/default/files/kvaliteet/Kokkuv%C3%B5te.pdf> (vaadatud 07.04.2024).
- [4] Mark L. Levy. „The national review of asthma deaths: what did we learn and what needs to change?“ *Breathe (Sheff)* 11.1 (2015), lk. 14–24. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4487386/> (vaadatud 07.04.2024).
- [5] „Astma“. *Eesti Kopsuliidu kodulehekülg* (2021). URL: <https://www.kopsuliit.ee/haigused/astma/> (vaadatud 07.04.2024).
- [6] „Chronic respiratory diseases: asthma“. *Maaailma Terviseorganisatsiooni kodulehekülg* (2021). URL: <https://www.who.int/news-room/questions-and-answers/item/chronic-respiratory-diseases-asthma> (vaadatud 07.04.2024).
- [7] *EXPANSE projekti kodulehekülg*. 2024. URL: <https://expanseproject.eu/> (vaadatud 20.04.2024).



- [8] „Basic Information about NO<sub>2</sub>“. *United States Environmental Protection Agency kodulehekülg* (2023). URL: <https://www.epa.gov/no2-pollution/basic-information-about-no2#Effects> (vaadatud 20.04.2024).
- [9] *Keskkonnaministri 27.12 määrus nr 75 „Õhukvaliteedi piir- ja sihtväärtused, õhukvaliteedi muud piirnormid ning õhukvaliteedi hindamispiirid“*. 2016. URL: <https://www.riigiteataja.ee/akt/106032019012> (vaadatud 08.04.2024).
- [10] „What are the WHO Air quality guidelines?“ *Maailma Terviseorganisatsiooni kodulehekülg* (2021). URL: <https://www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines> (vaadatud 08.04.2024).
- [11] „Ground-level Ozone Basics“. *United States Environmental Protection Agency kodulehekülg* (2023). URL: <https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics#effects> (vaadatud 20.04.2024).
- [12] „Health and Environmental Effects of Particulate Matter (PM)“. *United States Environmental Protection Agency kodulehekülg* (2023). URL: <https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm> (vaadatud 08.04.2024).
- [13] Antonio Voce *et al.* „Europe’s pollution divide: see how your area compares“. *The Guardian* (2023). URL: <https://www.theguardian.com/environment/ng-interactive/2023/sep/20/europes-pollution-divide-see-how-your-area-compares> (vaadatud 19.04.2024).

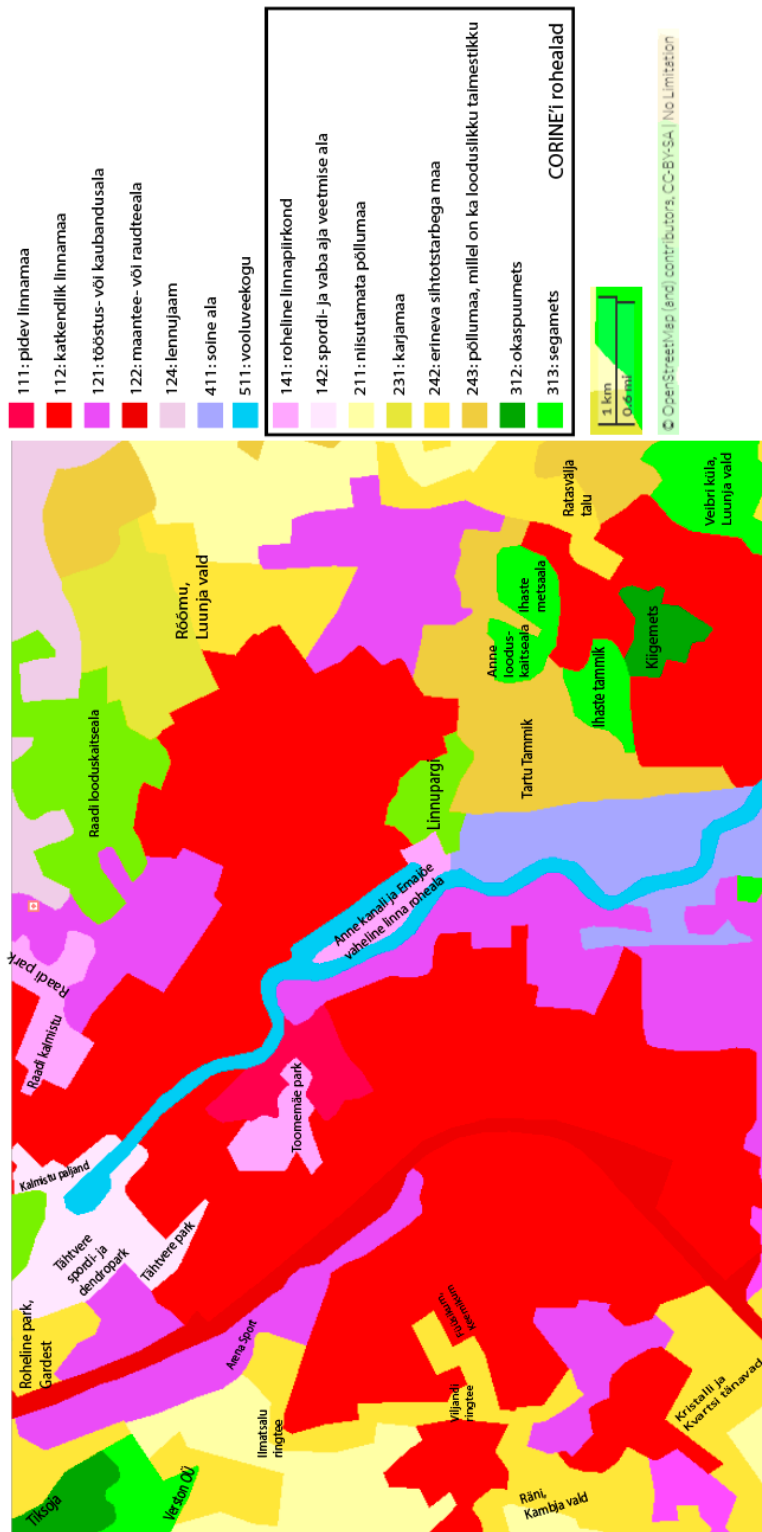
- [14] „CORINE Land Cover“. *The Copernicus Land Monitoring Service kodulehekülg* (2024). URL: <https://land.copernicus.eu/en/products/corine-land-cover?tab=main> (vaadatud 19.04.2024).
- [15] *Description of Environmental variables available through Exposome Maps*. Versioon 1.0. 2023. URL: <https://surfdrive.surf.nl/files/index.php/s/uqUORDrd428H2F9> (vaadatud 19.04.2024).
- [16] Ireneous N. Soyiri ja Ian Alcock. „Green spaces could reduce asthma admissions“. *The Lancet Respiratory Medicine* 6.1 (2018), lk. e1. URL: <https://www.sciencedirect.com/science/article/pii/S2213260017304411> (vaadatud 14.05.2024).
- [17] Elisa T. Lee ja John Wenyu Wang. *Statistical methods for survival data analysis*. Fourth edition. John Wiley & Sons, Inc., 2013.
- [18] Zach Bobbitt. „How to Use the scale() Function in R (With Examples)“. *Statology* (2021). URL: <https://www.statology.org/scale-function-in-r/> (vaadatud 16.04.2024).
- [19] Richard A. Johnson ja Dean W. Wichern. *Applied Multivariate Statistical Analysis*. Third edition. Prentice-Hall International, Inc., 1992.
- [20] Tomas M.L. Eagan *et al.* „The effect of educational level on the incidence of asthma and respiratory symptoms“. *Respiratory Medicine* 98.8 (2004), lk. 730–736. URL: <https://www.sciencedirect.com/science/article/pii/S0954611104000575> (vaadatud 20.03.2024).
- [21] Erick Forno ja Juan C. Celedón. „The effect of obesity, weight gain, and weight loss on asthma inception and control.“ *Current Opinion in Allergy and Clinical Immunology* 17.2 (2017), lk. 123–130. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5545117/> (vaadatud 09.05.2024).

- [22] Hubaida Fuseini ja Dawn C. Newcomb. „Mechanisms Driving Gender Differences in Asthma.“ *Current allergy and asthma reports* 17.3 (2017), lk. 19. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5629917/> (vaadatud 10.05.2024).
- [23] Kewu Huang *et al.* „Prevalence, risk factors, and management of asthma in China: a national cross-sectional study“. *Lancet* 394.10196 (2019), lk. 407–418. URL: <https://spiral.imperial.ac.uk/handle/10044/1/71090> (vaadatud 10.05.2024).
- [24] Hamood Ur-Rehman Malik *et al.* „Minimal difference in the prevalence of asthma in the urban and rural environment“. *Clin Med Insights Pediatr* 6 (2012), lk. 33–39. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3620776/> (vaadatud 14.05.2024).
- [25] Abdulbari Bener *et al.* „Prevalence of asthma and wheeze in two different climatic areas of Saudi Arabia“. *The Indian journal of chest diseases allied sciences* 35.1 (1993), lk. 9–15. URL: <https://pubmed.ncbi.nlm.nih.gov/8225434/> (vaadatud 14.05.2024).
- [26] C. Asam *et al.* „Tree pollen allergens—an update from a molecular perspective“. *Allergy* 70.10 (2015), lk. 1201–1211. URL: <https://onlinelibrary.wiley.com/doi/full/10.1111/all.12696> (vaadatud 13.05.2024).
- [27] „Õietolmu seire“. *Eesti Allergialiidu kodulehekülg* (2024). URL: <https://allergialiit.ee/oietolmu-seire/> (vaadatud 13.05.2024).
- [28] Louise Duquesne *et al.* „The influence of urban trees and total vegetation on asthma development in children“. *Environmental Epidemiology* 7.6 (2023). URL: [https://journals.lww.com/environepidem/fulltext/2023/12000/the\\_influence\\_of\\_urban\\_trees\\_and\\_total\\_vegetation.8.aspx](https://journals.lww.com/environepidem/fulltext/2023/12000/the_influence_of_urban_trees_and_total_vegetation.8.aspx) (vaadatud 14.05.2024).

- [29] Eija Parmes *et al.*, „Influence of residential land cover on childhood allergic and respiratory symptoms and diseases: Evidence from 9 European cohorts“. *Environmental Research* 183 (2020), lk. 108953. URL: <https://www.sciencedirect.com/science/article/pii/S0013935119307509> (vaadatud 14.05.2024).

Lisad

Lisa 1 - Tartu linna CORINE'i alade kaart



## Lisa 2 - Andmestiku üldtunnused

Tabel 8: Andmestiku üldtunnused

Tunnuse lühend	Antud väärtusi	Kirjeldus	Puuduvaid väärtusi
<i>astma_diagnoos</i> <i>diagnoos_kp</i> <b>astma</b> <i>esma_kp</i>	0 kuni mitu 0 kuni mitu 1 1	Astma diagnoosi tähis ja saamise kuupäev. Kui doonor on astmaatik (vt ptk 1), siis on tunnuse <i>astma</i> väärtus 1 (vastasel juhul 0) ning märgitud on esimese diagnoosi saamise kuupäev.	
<i>surm</i>	1	Kui doonor on surnud enne 2023. aastat, siis on antud surma kuupäev. Vastasel juhul väärtus puudub.	
<b>aeg</b>	1	Mitu aastat kulus 2013. a-st, kuni doonor sai astma või suri, maksimum väärtus on 10 a.	
<b>sugu</b>	1	Väärtuseks mees (1) või naine (2).	0
<i>sunniaasta</i> <b>vanus</b>	1	Doonori sünniaasta. Vanuse väärtus on 2023. a ja sünniaasta vahe aastates.	0 0
<i>liitumisaeg</i> <b>liitumisperiod</b>	1	Kuupäev, millal doonor liitus Geenivaramuga. <i>liitumisperiod</i> on 1, kui <i>liitumisaeg</i> on alates 2018. a-st (vastasel juhul 0).	0 0

Tunnuse lühend	Antud väärtusi	Kirjeldus	Puuduvaid väärtusi
<i>bmi</i>	1 kuni 4	Küsimustikust saadud kaalu ja pikkuse põhjal arvutatud kehamassiindeks.	46 473
<i>suits_staatus</i> <i>suits</i>	1 kuni 4 1 kuni 4	Küsimustikust saadud suitsetamise staatus: mitte kunagi, praegune või endine. Tunnusel <i>suits</i> on väärtus 0, kui doonor pole mitte kunagi suitsetanud (vastasel juhul 0).	45 882 45 882
<i>haridus</i>	1 kuni 4	Küsimustiku täitmisel vastatud kõige kõrgem kehtiv hariduse tase. Tasemeid on 9: ei tea, puudub, alg, põhi, kesk, kutse, rakenduslikkõrg, kõrg, teaduskraad.	52 324



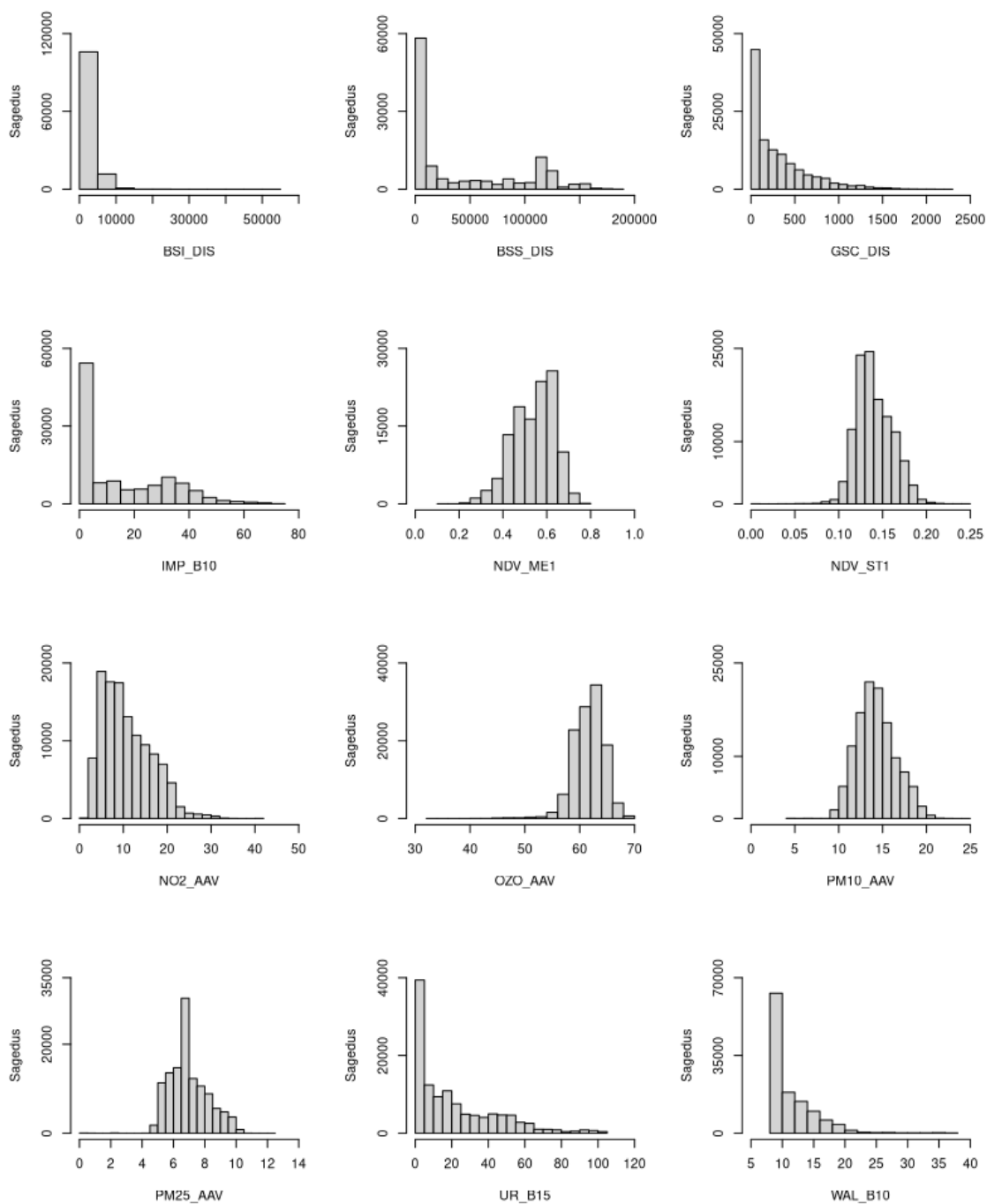
## Lisa 3 - Andmestiku eksposoomitunnused

Tabel 9: Andmestiku eksposoomitunnused

Tunnuse lühend ingl	Antud aastad	Kirjeldus	Puuduvaid väärtusi
<b>BSI_DIS</b>	<b>2013</b>	Kaugus lähimast siseveekogust meetrites.	3809
<b>BSS_DIS</b>	<b>2013</b>	Kaugus lähimast merepiirist meetrites.	3809
<i>BSW_DIS</i>	2013	Kaugus lähimast veekogust meetrites.	3809
<b>GSC_DIS</b>	2000, 2006, <b>2012</b> , 2018	Kaugus rohealadele CORINE'i põhjal.	3825
<i>GSU_DIS</i>	2006, 2012, 2018	Kaugus rohealadele <i>Urban</i> atlase põhjal.	70 076
<i>IMP_B03</i>	kõigil 2006,	Kõvade pindade, näiteks	3809
<i>IMP_B05</i>	2009, 2012,	kõvakattega teed, parklad,	3810
<b>IMP_B10</b>	<b>2015</b> , 2018	katused, protsent kas 300, 500 või 1000 m raadiuses maaalas.	3907
<i>LAN_B03</i>	kõigil 2000,	Tehisvalguse hulk öösel kas	30 282
<i>LAN_B05</i>	2005, 2010,	300, 500 või 1000 m	19 432
<i>LAN_B10</i>	2015, 2020	raadiuses.	35 416
<i>MVI_MD3</i>	kõigil 2000,	Rohelust ilmestav MV indeksi	4442
<i>MVI_MD5</i>	2005, 2010,	aastane mediaan kas 300, 500	4225
<i>MVI_MD1</i>	2015, 2020	või 1000 m raadiuses.	4426
<i>MVI_ME3</i>	kõigil 2000,	Rohelust ilmestav MV indeksi	4442
<i>MVI_ME5</i>	2005, 2010,	aastane keskmine kas 300,	4225
<i>MVI_ME1</i>	2015, 2020	500 või 1000 m raadiuses.	4426
<i>MVI_ST3</i>	kõigil 2000,	Rohelust ilmestav MV	4454
<i>MVI_ST5</i>	2005, 2010,	indeksi aasta keskmine	4254
<i>MVI_ST1</i>	2015, 2020	standardhälve kas 300, 500 või 1000 m raadiuses.	4460

<b>Tunnuse lühend ingl</b>	<b>Antud aastad</b>	<b>Kirjeldus</b>	<b>Puuduvaid väärtusi</b>
<i>NDV_MD3</i> <i>NDV_MD5</i> <i>NDV_MD1</i>	kõigil 2000, 2005, 2010, 2015, 2020	Rohelust ilmestav NDV indeksi aastane mediaan kas 300, 500 või 1000 m raadiuses.	4442 4225 4431
<i>NDV_ME3</i> <i>NDV_ME5</i> <b><i>NDV_ME1</i></b>	kõigil 2000, 2005, 2010, <b>2015</b> , 2020	Rohelust ilmestav NDV indeksi aastane keskmine kas 300, 500 või 1000 m raadiuses.	4442 4225 4431
<i>NDV_ST3</i> <i>NDV_ST5</i> <b><i>NDV_ST1</i></b>	kõigil 2000, 2005, 2010, <b>2015</b> , 2020	Rohelust ilmestav NDV indeksi aasta keskmine standardhälve kas 300, 500 või 1000 m raadiuses.	4454 4254 4465
<b><i>NO2_AAV</i></b>	2000, ..., <b>2013</b> , ..., 2019	Aasta keskmine lämmastikdioksiidi kontsentratsioon õhus ühikus $\mu g/m^3$ .	3827
<b><i>OZO_AAV</i></b>	2000, ..., <b>2013</b> , ..., 2019	Aasta keskmine osooni kontsentratsioon õhus ühikus $\mu g/m^3$ .	3827
<b><i>PM10_AAV</i></b>	2000, ..., <b>2013</b> , ..., 2019	Aasta keskmine peenete osakeste kontsentratsioon õhus ühikus $\mu g/m^3$ .	3827
<b><i>PM25_AAV</i></b>	2000, ..., <b>2013</b> , ..., 2019	Aasta keskmine ülipeenete osakeste kontsentratsioon õhus ühikus $\mu g/m^3$ .	3827
<b><i>UR_B15</i></b>	<b>2015</b>	Linnakeskkonna üldkvaliteeti hindav indeks.	3809
<i>WAL_B03</i> <i>WAL_B05</i> <b><i>WAL_B10</i></b>	kõigil <b>2020</b>	Jalutamiseks sobivuse indeks kas 300, 500 või 1000 m raadiuses.	4007 4058 4290

## Lisa 4 - Valitud eksposoomitunnuste histogrammid



## **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Hanna-Maria Kukk,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Väliskeskonna mõju astmariskile“, mille juhendajad on Krista Fischer ja Jaanika Kronberg, 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.

Hanna-Maria Kukk

15.05.2024