

TARTU ÜLIKOOL

LOODUS- JA TÄPPISTEADUSTE VALDKOND

MATEMAATIKA JA STATISTIKA INSTITUUT

Emma Räim

**Keskkonnategurite mõju komorbiidsuse
võrgustikele TÕ Eesti geenivaramu andmete
põhjal**

Matemaatiline statistika

Bakalaureusetöö (9 EAP)

Juhendajad: PhD Jaanika Kronberg, Phd Krista Fischer

TARTU 2024

KESKKONNATEGURITE MÕJU KOMORBIIDSUSE VÕRGUSTIKELE TÕ EESTI GEENIVARAMU ANDMETE PÕHJAL

Bakalaureusetöö

Emma Räim

Lühikokkuvõte

Mittenakkuslikud haigused, sageli tuntud krooniliste haigustena, tulenevad geneetiliste, füsioloogiliste, keskkonna- ja käitumiskomponentide koosmõjust. Ülemaailmne haiguste koorem on nihkunud nakkushaigustest mittenakkuslike haiguste poole. Haiguste geneetilist riski muuta ei saa, samuti nõuab elustiili muutmine motivatsiooni. Väliskeskkonna kui haigusi mõjutava komponendi muutused võivad potentsiaalselt vähendada haiguskoormat, rõhutades vajadust uurida, missugused keskkonnakomponendid on haigustega seotud. Võrgustikepõhine analüüs võimaldab hinnata keskkonnategurite mõju paljudele haigustele. Töö eesmärk on uurida keskkonnategurite mõju komorbiidsuse ehk mitme haiguse koosesinemise võrgustikele, kasutades selleks Eesti geenivaramu andmeid. Töös leiti, et suitsetamine ja haridus avaldavad võrgustikele oodatavat mõju: suitsetamine suurendab komorbiidsust, haridus vähendab. See-eest ülipeente osakeste mõju suurendas mõnede haiguste komorbiidsuskoores ja teistel vähendas.

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

Märksõnad: Komorbiidsus, eksposoom, Eesti geenivaramu.

**ENVIRONMENTAL INFLUENCES ON COMORBIDITY
NETWORKS: ANALYSIS OF ESTONIAN BIOBANK DATA**

Bachelor thesis

Emma Räm

Abstract

Non-communicable diseases, often referred to as chronic diseases, result from a combination of genetic, physiological, environmental, and behavioral factors. The global burden of disease has shifted from infectious diseases towards non-communicable diseases. Genetic risk for diseases cannot be altered, and lifestyle changes require motivation. Changes in the external environment, as a component influencing diseases, could potentially reduce the disease burden, highlighting the need to investigate which environmental components are associated with diseases. Network-based analysis enables the assessment of the impact of the environmental factors on many diseases. The aim of this study is to investigate the influence of environmental factors on comorbidity networks, i.e., networks illustrating the co-occurrence of multiple diseases, using data from the Estonian Biobank. The study revealed that smoking and education have the anticipated influence on networks: smoking increases comorbidity, while education decreases it. However, the concentration of fine particulate matter heightened comorbidity scores for certain diseases and lowered them for others.

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

Key Words: Comorbidity, exposome, Estonian Biobank.

Sisukord

Sissejuhatus	5
1 Bioloogiline taust	7
1.1 Komorbiidsus	8
1.2 Vaadeldavad keskkonnategurid	9
1.2.1 Suitsetamine ja haridus	9
1.2.2 Ülipeenete osakeste kontsentratsioon PM _{2,5}	10
2 Metoodika	12
2.1 Komorbiidsuse skoor	12
2.2 Fisheri statistiku transformatsioon	13
2.3 Graafid	15
2.3.1 Graafid bioloogias	16
2.3.2 Graafide klasterdamine	16
3 Andmeanalüüs	18
3.1 Andmed	18
3.1.1 Komorbiidsuse tabel	18
3.1.2 Valitud haigused	19
3.2 Komorbiidsuse tabel kogu andmestiku peale	20
3.3 Komorbiidsuse tabelite võrdlus haiguste paaride kaupa	21
3.3.1 Kahe mittekattuva juhuslikult valitud hulga komorbiidsus- kordajad	21
3.3.2 Suitsetajate ja mittedsuitsetajate hulkade komorbiidsuskordajad	22

3.3.3	Kõrgharidusega ja kõrghariduseta hulkade komorbiidsuskordajad	24
3.3.4	Ülipeente osakeste kontsentratiooni järgi grupeeritud hulkade komorbiidsuskordajad	25
3.4	Komorbiidsusvõrgustikud	26
3.4.1	Suitsetajate ja mittesuitsetajate hulkade komorbiidsusvõrgustikud	27
3.4.2	Kõrgharidusega ja kõrghariduseta hulkade komorbiidsusvõrgustikud	30
3.4.3	Ülipeente osakeste kontsentratiooni järgi grupeeritud hulkade komorbiidsusvõrgustikud	31
3.4.4	Kahe mittekattuva juhuslikult valitud hulga komorbiidsusvõrgustikud	34
3.5	Komorbiidsuskordajate erinevuste võrdlus tegurite kaupa	36
	Kokkuvõte	38
	Kasutatud allikad	40
	Lisa 1. Võrgustik	47
	Lisa 2. Komorbiidsuskordajate erinevuste võrdlus	49

Sissejuhatus

Mittentakuslikud haigused, tihtipeale tuntud krooniliste haigustena, on pikaajsed haigused, mis on tulemus geneetiliste, füsioloogiliste, keskkonna- ja käitumistegurite kooslusest (World Health Organization, 2023). Kuigi enim on mittentakuslikest haigustest mõjutatud inimesed madala ja keskmise sissetulekuga riikidest (World Health Organization, 2023), siis on haiguste koorem ülemaailmselt nihkunud nakkuslikest haigustest mittentakuslike suunas (Münzel, Sørensen, Hahad, Nieuwenhuijsen ja Daiber, 2023).

On leitud, et keskkonnal on arvatust palju suurem mõju ülemaailmsele haiguskoormale (Münzel, Sørensen, Hahad, Nieuwenhuijsen ja Daiber, 2023), mis on tõstatanud vajaduse keskkonnateguritel põhinevale lähenemisele, mis aitaks leida seoseid haiguste ja keskkonnategurite vahel. Üheks meetodiks on välja pakutud võrgustike loomist eksposoomi¹ ja haiguste vaheliste seoste analüüsimiseks. Varasemalt on uuritud seoseid keskkonnategurite ja üksikute haiguste vahel (Fan, Li, Fan, Bai ja Yang, 2015; Hamra, Laden, Cohen, Raaschou-Nielsen, Brauer ja Loomis, 2015), kuid paljude haiguste võrgustikuanalüüs võiks näidata, mis haigusi oleks huvitav detailsemalt edasi uurida.

Käesolevas töös kasutatakse komorbiidsusepõhist lähenemist. Kaks haigust on komorbiidsed², kui neil on ühine jagatud mehhanism, näiteks bioloogiline rada (Valderas, Starfield, Sibbald, Salisbury ja Roland, 2009). Muutused jagatud bioloogilises rajas võiks kajastuda ka haiguste komorbiidsuse skooris. Selle idee testimiseks valiti kaks vastandlikku keskkonnategurit ning üks uudne eksposoomitegur. Töö eesmärk on uurida antud kolme teguri mõju komorbiidsuse võrgustikele. Komorbiidsuse võrgustikud arvutatakse, kasutades Eesti geenivaramu andmeid (Leitsalu et al., 2014).

¹Kõik tegurid, mis inimest mõjutavad (füüsiline, sotsiaalne ja väliskeskkond, enda valikud ja palju muud)

²Üksikisikul vähemalt kahe haiguse samaaegne esinemine

Esimeses peatükis kirjeldatakse täpsemalt lahti uuritav probleem ning tutvustatakse tööse kaasatud keskkonnategureid. Teises peatükis on kirjeldatud töös kasutatavat meetodikat. Kolmandas peatükis on tutvustatud andmeid ning välja toodud analüüsitud hulkade vahelised komorbiidsuse seosed ja komorbiidsuse erinevuste võrgustikud.

Autor tänab bakalaureusetöö juhendajaid Jaanika Kronbergi ja Krista Fischerit nende olulise toetuse ning TÜ Eesti geenivaramut andmetega töötamise võimaluse eest.

1 Bioloogiline taust

Surmade ja puuete põhjused võib kategoriseerida kolmeks suureks grupiks: nakkushaigusteks, mittenakkuslikeks haigusteks ning vigastusteks (World Health Organization, 2020). On kasvavaid tõendeid, et haiguste koorem on ülemaailmselt nihkunud nakkuslikest haigustest mittenakkuslike suunas (Münzel, Sørensen, Hahad, Nieuwenhuijsen ja Daiber, 2023) ja seda mitte ainult kõrge, vaid ka keskmise ning madala sissetulekutega riikides (Goryakin, Rocco ja Suhrcke, 2017). Vastavalt Maailma Terviseorganisatsiooni (edaspidi WHO) hinnangutele olid mittenakkuslikud haigused 2019. aastal vastutavaks 73,6% maailma surmadest, millest omakorda ligikaudu 41,8% loeti nii-öelda enneaegseteks ehk enne 70. eluaastat esinevateks (World Health Organization, 2020).

Mittenakkuslikest haigustest moodustasid kardiometaboolsed haigused 43,8%, mis oli kogu surmade osakaalust 32,2%. Kardiometaboolsetest haigustest suurima osa kogu surmade osakaalule andsid südame isheemiatõved 16,0% ning insult 11,2%. Küll väiksema, kuid märgatava osa ehk 7,5% surmadest moodustasid kopsuhaigused, millest krooniline obstruktiivne kopsuhaigus moodustas kopsuhaigustest 77,9% ehk 5,8% kogu surmade osakaalust. (World Health Organization, 2020)

Inimene ei saa muuta oma geneetilisi riske, kuid haigusriske saab mõjutada elustiilivalikutega. Näiteks on suitsetamise ja füüsilise aktiivsuse mõjud seostatud mittenakkuslike haigustega (Tolonen, Reinikainen, Zhou, Härkänen, Männistö, Jousilahi, Paalanen, Lundqvist ja Laatikainen, 2022). Samuti, hea südame-veresoonkonna tervis keskeas võib vähendada krooniliste haiguste ja suremuse riske (Corlin, Short, Vasan ja Xanthakis, 2020). Niisiis on oluline roll elustiilivalikutel.

Lisaks selgitab geneetiline varieeruvus paljude keerulisemate haiguste, nagu vähk, teist tüüpi diabeet, südame-veresoonkonna ja kopsuhaigused, puhul ainult tagasihoidliku osa haigusriskist (Münzel, Sørensen, Hahad, Nieuwenhuijsen ja Daiber, 2023). Näiteks Lääne-Euroopa ühemunarakukaksikute andmete põhjal läbiviidud uuringus selgus, et ainult 16,4% vähkide ja südamehaiguste surmadest võib omis-

tada geneetiliselt seotud teguritele (Rappaport, 2016). Niisiis, võib keskkonnal olla arvatust palju suurem mõju ülemaailmsele haiguskoormale (Münzel, Sørensen, Hahad, Nieuwenhuijsen ja Daiber, 2023).

See on tõstatanud vajaduse uuele eksposoomipõhisele lähenemisele, mis suudaks leida seoseid haiguste ja keskkonnategurite vahel. Eksposoomiga mõõdetakse kõiki üksikisiku eluajal tekkinud kokkupuuteid ning nende mõju tervisele (National Institute for Occupational Safety and Health, 2022). Haiguse ja eksposoomi vahelisi seoseid on analüüsitud nii südame-veresoonkonna kui ka kopsuhaiguste puhul (Münzel, Sørensen, Hahad, Nieuwenhuijsen ja Daiber, 2023; Wheelock ja Rappaport, 2020).

Eksposoomi ja haiguste seoste mõistmine võib olla keeruline, kuna eksposoom on kompleksne ja arvatakse koosnevat mitmest omavahel sõltuvast individuaalsest kokkupuutest (Patel ja Manrai, 2015). Võrgustike loomine haiguste ja geenide vaheliste seoste hindamiseks on osutunud efektiivseks meetodiks keerukate interaktsioonide mõistmiseks (Goh, Cusick, Valle, Childs, Vidal ja Barabási, 2007). Samuti on rakendatud võrgustike loomist mitmete molekulaarsete interaktsioonide analüüsimiseks (Barabási ja Oltvai, 2004). Niisiis, võib osutada võrgustike meetod tõhusaks ka eksposoomi ja haiguste vaheliste seoste uurimiseks.

1.1 Komorbiidsus

Komorbiidsus on üksikisikul vähemalt kahe haiguse samaaegne esinemine. Antud töös vaadatakse komorbiidsust kui kahe haiguse koosinemist. Komorbiidsust on seostatud halvemate tervisetulemuste, keerulisema kliinilise juhtimise ja suuremate tervishoiukuludega (Valderas, Starfield, Sibbald, Salisbury ja Roland, 2009). Komorbiidsuse esinemine tekib jagatud bioloogiliste radade tõttu (Valderas, Starfield, Sibbald, Salisbury ja Roland, 2009). Kui nüüd mingi väline faktor mõjutab mingit kindlat rada, siis kõik sellega seotud haigused võivad olla mõjutatud. See omakorda võib tuua kaasa näiteks suurenenud komorbiidsuse.

Eksposoomipõhiste komorbiidsusvõrgustike analüüs võib aidata kaasa eksposoomi ja haiguste vaheliste seoste tuvastamisele. Nimelt, kui fikseerida mõni väline tegur ning kasutada seda tingimusena hulkadesse jagamisel, mille põhjal komorbiidsused arvutatakse, siis eeldus on, et välise teguri mõju korral komorbiidsused hulkades erinevad. See võimaldab eksposoomi mõjude analüüsimist lihtsustatud viisil.

Näiteks viidi läbi uuring *Phenotypic Disease Network* andmete põhjal, kus leiti komorbiidsuseid 30-ne miljoni isiku andmete abil. Selle põhjal loodud komorbiidsuse võrgustik näitas, et haiguse progresseerumist saab esindada ja uurida võrgustikumeetodite abil, potentsiaaliga mõista inimhaiguste päritolu ja evolutsiooni varasemast paremini. (Hidalgo, Blumm, Barabási ja Christakis, 2009)

Lisaks on uuritud võrgustikumeetodi valiidsust ning tulemused on toetanud meetodi kehtivust. Küll aga on leitud, et meetodiga saadud tulemused muutuvad vähem usaldusväärseks, kui keskenduda üksikutele haigustele eraldi. (Sadegh *et al.*, 2023)

1.2 Vaadeldavad keskkonnategurid

Käesolevasse töösse kaasati kolm keskkonnategurit: suitsetamine, haridus ja ülipeente osakeste kontsentratsioon. Tunnuste haridus ja suitsetamine kohta on kogutud andmed läbi Geenivaramu küsimustiku. Peenosakeste kontsentratsioon on eksposoomitunnus keskkonnaandmetest, mis ühendati EXPANSE projekti raames geenidoonorite andmetega ning mida varem ei olnud võimalik geenidoonoritel uurida (Vlaanderen *et al.*, 2021).

1.2.1 Suitsetamine ja haridus

2019. aastal läbiviidud uuringus *Global Burden of Disease Study 2019* oli tubaka suitsetamine kõrge süstoolse vererõhu järel kõige levinum surma põhjustav riskitegur. Lisaks leiti, et mõned riskikäitumised on tugevalt korreleeritud riikide sotsiaalse ja majandusliku arenguga. Üheks arengule kaasaaitavaks teguriks on kõrg-

haridus. (Murray et al., 2020)

Uuringutest on leitud, et suitsetamine suurendab krooniliste mittenakkuslike haiguste, eriti südame-veresoonkonna haiguste, globaalset haiguskoormat (Münzel, Sørensen, Hahad, Nieuwenhuijsen ja Daiber, 2023). Hariduse puhul on näiteks täheldatud, et piiratud haridustase on seotud diabeedi levimuse suurenemisega (Tremblay ja Hamet, 2019, Wang, Stewart, McDermott, Kazanjian, Vissandjee, DesMeules, Groh ja Morrison, 2012). Niisiis eeldas autor analüüsi käigus näha esimese teguri puhul suurendavat ja teise teguri puhul vähendavat mõju komorbiidsusele.

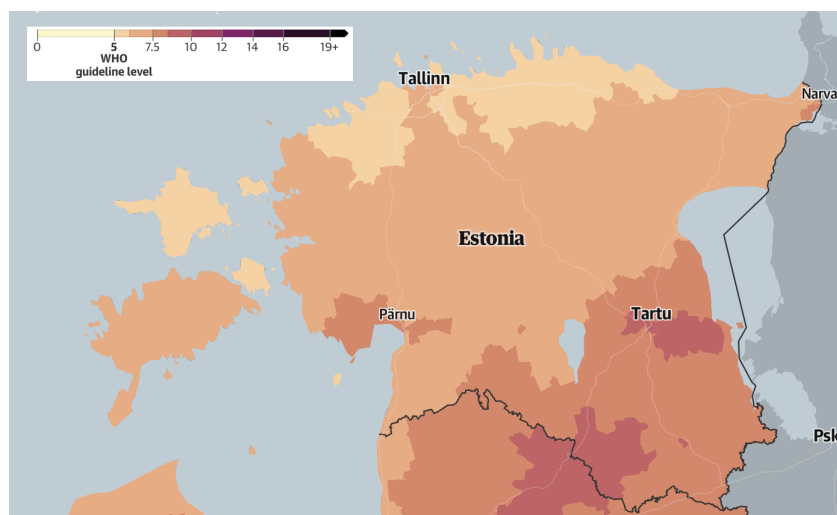
1.2.2 Ülipeenete osakeste kontsentratsioon PM_{2,5}

PM (ingl *Particulate Matter*) on õhus leiduvad peened osakesed, mis koosnevad mitmete keemiliste ainete segust. PM_{2,5} viitab ülipeentele osakestele, mille läbimõõt on alla 2,5 μm . (California Air Resources Board, 2019)

Uuringus *Global Burden of Disease Study 2019* leiti, et aastate 2010 – 2019 jooksul oli üheks suurima kasvuga riskiteguriks õhus leiduvate osakeste saaste (Murray et al., 2020). Lisaks oli 2020. aastal 96% linnapiirkondades elavate eurooplaste elukohta PM_{2,5} tase üle lubatud WHO kehtestatud piiri 5 $\mu\text{g}/\text{m}^3$ (European Environment Agency, 2022).

PM_{2,5} mõju on küll uuritud Tallinnas seoses suremuse ja haiglaravi vajadusega südame-veresoonkonna ning hingamisteede haiguste korral, kus uuritud valimi suuruseks oli 390 000, kuid puudub võimalus näiteks paljude erinevate haiguste paralleelseks analüüsimiseks (Orru, Teinemaa, Lai, Tamm, Kaasik, Kimmel, Kangur, Merisalu ja Forsberg, 2009).

Jooniselt 1 on näha, kuidas 2019. aasta ülipeente osakeste keskmine on kõikide piirkondade puhul üle WHO normi. Kõrgeima tasemega on Valga ja Tartu ümbrus, kus tase jääb vahemikku 8,5 – 10 $\mu\text{g}/\text{m}^3$. Madalaima taseme vahemikuga 5 – 6 $\mu\text{g}/\text{m}^3$ on suur osa Põhja-Eestist ning mitmed saared näiteks nagu Hiiumaa.



Joonis 1: Eesti 2019. aasta õhureostuse kontsentratsiooni PM_{2,5} kaart (Voce, Anna, Ellen ja Pamela, [2023](#))

2 Metoodika

2.1 Komorbiidsuse skoor

Olgu D1 ja D2 kaks erinevat haigust ning avaldugu nendevaheline sõltuvustabel järgmiselt:

Tabel 1: Kahe haiguse vaheline sõltuvustabel

D2	D1		Summa
	Esines	Ei esinenud	
Esines	a	b	$a + b$
Ei esinenud	c	d	$c + d$
Summa	$a + c$	$b + d$	$a + b + c + d$

Siis haiguste D1 ja D2 vahelise komorbiidsuse kordaja saame järgmiselt:

$$\phi = \frac{ad - bc}{\sqrt{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}}, \quad (1)$$

mis ei ole defineeritud, kui $c = d = 0$ või $b = d = 0$ ning mis saavutab väärtuseid vahemikust $[-1,1]$. (Hubálek, 1982)

Teisendame ϕ eelmainitud artiklis (Hidalgo, Blumm, Barabási ja Christakis, 2009) kasutatud kujule:

$$\begin{aligned} \phi &= \frac{a^2 + ab + ac + ad - a^2 - ac - ab - bc}{\sqrt{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}} \\ &= \frac{a \cdot (a+b+c+d) - (a+b) \cdot (a+c)}{\sqrt{(a+b) \cdot (a+c) \cdot (a+b+c+d-a-b) \cdot (a+b+c+d-a-c)}} \\ &= \frac{C_{ij} \cdot N - I_i \cdot I_j}{\sqrt{I_i \cdot I_j \cdot (N - I_i) \cdot (N - I_j)}}, \end{aligned} \quad (2)$$

kus $N := a + b + c + d$ on valimi maht; $C_{ij} := a$ inimeste arv, kes olid samaaegselt

diagnoositud haigustega i ja j ; $I_i := a + b$ haiguse i intsidentide arv ning $I_j := a + c$ haiguse j intsidentide arv.

2.2 Fisheri statistiku transformatsioon

Kahe tunnuse, X ja Y vahelise lineaarse seose tugevust hinnatakse enamasti Pearsoni korrelatsioonikordaja abil. Olgu meil N kahe tunnuse väärtuste paari (x_i, y_i) , $i = 1, \dots, N$, siis Pearsoni korrelatsioonikordaja avaldub valemiga:

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}. \quad (3)$$

Järgnev põhineb Dennis E. Hinkle'i raamatul "Applied statistics for the behavioral sciences", kui ei ole öeldud teisiti (Hinkle DE, 1988). Korrelatsioonikordajad transformeeritakse kasutades Fisheri statistiku z transformatsiooni:

$$z_r = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right),$$

ja korrelatsioonikordajate erinevuse standardvea saame:

$$se_{z_{r_1} - z_{r_2}} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}.$$

Fisheri statistik z kahe sõltumatu korrelatsioonikordaja erinevuse statistilise olulisuse leidmiseks avaldub kujul:

$$z = \frac{z_{r_1} - z_{r_2}}{se_{z_{r_1} - z_{r_2}}}$$

Kui (X, Y) on ühise kahemõõtmelise normaaljaotusega või valimi suurus n on suur, siis Fisheri transformeeritud z_r korrelatsioonikordajaga r ning valimi suurusega n on normaaljaotusega.

Selleks, et Fisheri statistiku transformatsiooni rakendada komorbiidsuskordajale ϕ (valem 2), näitame, et

$$\phi = \frac{ad - bc}{\sqrt{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} = r.$$

Tõestus. Olgu valimis suurusega n sagedusmaatriks antud kujul

X/Y	1	0	Σ
1	a	b	$a + b$
0	c	d	$c + d$
Σ	$a + c$	$b + d$	n

Siis

$$\begin{aligned} a &= \sum_{i=1}^n x_i y_i \\ b &= \sum_{i=1}^n x_i - \sum_{i=1}^n x_i y_i = \sum_{i=1}^n x_i (1 - y_i) \\ c &= \sum_{i=1}^n y_i - \sum_{i=1}^n x_i y_i = \sum_{i=1}^n y_i (1 - x_i) \\ d &= \sum_{i=1}^n (1 - x_i)(1 - y_i) \end{aligned}$$

Seega saame, et

$$\begin{aligned} ad &= \sum_{i=1}^n x_i y_i \sum_{i=1}^n (1 - x_i)(1 - y_i) = \\ &= n \sum_{i=1}^n x_i y_i - n\bar{y} \sum_{i=1}^n x_i y_i - n\bar{x} \sum_{i=1}^n x_i y_i + \left(\sum_{i=1}^n x_i y_i \right)^2 \\ bc &= \sum_{i=1}^n x_i (1 - y_i) \sum_{i=1}^n y_i (1 - x_i) = \\ &= n^2 \bar{x} \bar{y} - n\bar{x} \sum_{i=1}^n x_i y_i - n\bar{y} \sum_{i=1}^n x_i y_i + \left(\sum_{i=1}^n x_i y_i \right)^2 \end{aligned}$$

ning

$$ad - bc = n \sum_{i=1}^n x_i y_i - n^2 \bar{x} \bar{y} = n \left(\sum_{i=1}^n x_i y_i - n\bar{x} \bar{y} \right)$$

Lisaks näeme, et

$$\begin{aligned}(a+b)(c+d)(a+c)(b+d) &= \sum_{i=1}^n x_i \sum_{i=1}^n (1-x_i) \sum_{i=1}^n y_i \sum_{i=1}^n (1-y_i) = \\ &= (n^2\bar{x} - (n\bar{x})^2) (n^2\bar{y} - (n\bar{y})^2) = n^4 (\bar{x} - \bar{x}^2) (\bar{y} - \bar{y}^2)\end{aligned}$$

Kuna tunnused X ja Y on binaarsed, siis

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \sum_{i=1}^n (y_i^2 - 2y_i\bar{y} + \bar{y}^2) = \\ &= n^2 (\bar{x} - \bar{x}^2) (\bar{y} - \bar{y}^2) \\ \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^n (x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x} \bar{y}) = \\ &= \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}\end{aligned}$$

Järelikult saame näidata, et

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{n\sqrt{(\bar{x} - \bar{x}^2)(\bar{y} - \bar{y}^2)}} = \frac{\mathcal{X}(\sum x_i y_i - n\bar{x}\bar{y})}{n^{\frac{1}{2}}\sqrt{(\bar{x} - \bar{x}^2)(\bar{y} - \bar{y}^2)}} = \phi.$$

□

2.3 Graafid

Olgu meil antud mingid kindlad objektid, kus igale neist seame tasandil vastavusse ühe kindla punkti. Olenevalt sellest, millist objektidevahelist seost soovime näidata, ühendame punkte omavahel joontega. Kirjeldatud skeemi nimetatakse graafiks, punkte graafi tippudeks ning jooni graafi servadeks. Eeldame, et ükskõik millist kahte tippu ühendab ülimalt lõplik arv servi. Tipuga ühenduses olevate servade arvu nimetatakse tipu astmeks. (Koit, 2022)

Graafid võivad olla, kas suunatud või suunamata. Suunatud graafide puhul on kahe tipu vaheline seos selgelt määratletud suunaga, suunamata graafi servadel on vastastikune seos. Näiteks, kui haigusel A esineb komorbiidsus haigusega B, siis

haigusel B esineb komorbiidsus haigusega A. (Barabási ja Oltvai, 2004)

2.3.1 Graafid bioloogias

Graafe on võimalik rakendada bioloogia ja meditsiini valdkonnas näiteks valgu või geeni funktsiooni määramisel, efektiivsete strateegiate väljatöötamiseks erinevate haiguste raviks või varajaseks diagnoosimiseks. (Pavlopoulos, Secrier, Moschopoulos, Soldatos, Kossida, Aerts, Schneider ja Bagos, 2011)

Käesolevas töös rakendatakse suunamata graafe haigustevahelise komorbiidsuse visualiseerimiseks, kus haigused on tippudeks ning kus serva paksuse ja värvi abil saab näidata tunnuste väärtuseid, näiteks seose tugevust või erinevust erinevates hulkades.

Võrgustike loomiseks kasutatakse tarkvara Cytoscape (Shannon, Markiel, Ozier, Baliga, Wang, Ramage, Amin, Schwikowski ja Ideker, 2003).

2.3.2 Graafide klasterdamine

Graafide tõlgendamiseks saab neid klasterdada. Graafi klasterdamise all mõeldakse graafi jagamist alamgraafideks, kus iga alamgraafi sees on tegemist tugevate seostega ja erinevate alamraafide vahel on seosed nõrgemad. Üheks klasterdamise meetodiks on algoritmide kasutamine, leides nendega graafidel tihedalt ühendatud piirkondi. Cytoscape'is on mitmeid erinevaid klasterdamise algoritme, üks nendest on bioloogias laialt kasutatav *Community Clustering* (GLay). GLay-l on võrreldes mitmete teiste algoritmidega suur eelis - klastrite arvu ei pea eelnevalt määrama. (Morris, Apeltsin, Newman, Baumbach, Wittkop, Su, Bader ja Ferrin, 2011)

Graafide klasterdamine on eriti tähtis suurte võrgustike korral, kuna võimaldab leida bioloogiliselt olulisi alamgraafe, mis võivad molekulaarsete mõõtmiste korral olla seotud näiteks erinevate bioloogiliste radadega. Käesolevas töös on võrgustikud väikesed ning klasterdamist kasutatakse visualiseerimise eesmärgil. Niisiis on GLay

üks valikutest, mis klasterdamiseks sobib, kuid võrgustike väiksuse tõttu saaks analoogse tulemuse ka mõne teise algoritmi rakendamisel. Tulevikutöös, rakendades sama analüüsi rohkemate haiguste puhul, võimaldab klasterdamine eristada erinevaid haigusgrupe, näiteks võiks südamehaigused olla seotud ühe klastriga.

3 Andmeanalüüs

3.1 Andmed

Analüüs viidi läbi 212 198 Eesti geenivaramuga liitunud geenidoonori andmete põhjal. Töö käigus andmeid isiku tasemel ei käsitletud, vaid uuriti erinevate haiguste paaride esinemiste arvu kogu andmestikus. Töös kasutatavad andmetabelid olid eelnevalt koostatud juhendaja Jaanika Kronbergi poolt. 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. Vaadeldavateks haigusteks valiti kardiometaaboolsed ja kopsuhaigused, mille analüüsimiseks on olemas eetikakomitee luba EXPANSE projektis (Vlaanderen *et al.*, 2021). Kokku oli analüüsitavaid haiguseid 23.

Analüüs viidi läbi kasutades tarkvara R.

3.1.1 Komorbiidsuse tabel

Geenidoonorid jagati vastavalt valitud keskkonnategurile kahte hulka ning mõlemas grupis toimus grupeerimine haiguste paaride kaupa eraldi. Näiteks jagati doonorid kaheks vastavalt sellele, kas isikul on kõrgharidus või mitte. Iga hulk sisaldas kõikide haiguste kombinatsioone kahe kaupa.

Tabelis 2 on toodud näide tabelite kujust mittesuitsetajate komorbiidsuse tabeli põhjal, kus D1 ja D2 tähistavad mingit kindlat kahte haigust ning N on kogu hulga suurus.

Olles viinud andmed kirjeldatud kujule, sai valemi 2 abil kõikidele haiguste paaridele arvutada komorbiidsuskordajad. Seetõttu on edaspidises analüüsikirjelduses kutsutud vaadeldavaid andmetabeleid komorbiidsuse tabeliteks. Lisaks haridusele

Tabel 2: andmetabelite kuju mittesuitsetajate näitel

D1	D2	D1_n	D2_n	D1D2_n	N
I20	I21	11939	1797	1270	104079
I20	I22	11939	1797	243	104079
I20	I25	11939	8895	5609	104079

rakendati samasugust meetodikat ka suitsetamisele ja aastakeskmisele ülipeente osakeste (PM_{2,5}) kontsentratsioonile. Võrdluseks keskkonnategurite komorbiidsuse tabelitele rakendati sama meetodikat ka kahele mittekattuvale juhuslikult valitud geenidonorite hulgale. Kokku rakendati edasist analüüsi kaheksale komorbiidsuse tabelile.

3.1.2 Valitud haigused

Kahe haiguse vahel esineb komorbiidsus, kui komorbiidsuse kordaja (valem 2) $\phi > 0$ (Hidalgo, Blumm, Barabási ja Christakis, 2009), niisiis kaasati edasisse analüüsi ainult need haiguste paarid, millel komorbiidsus esines. Lisaks eemaldati kõik need haiguste paarid, kus kummagi haiguse esinemine hulgas oli alla 100.

Kuna pidi säilima tingimus, et igas komorbiidsuse grupis analüüsitakse ühesuguseid haiguseid, siis jäeti kõikidesse komorbiidsuse tabelitesse alles ainult ühised haiguste paarid.

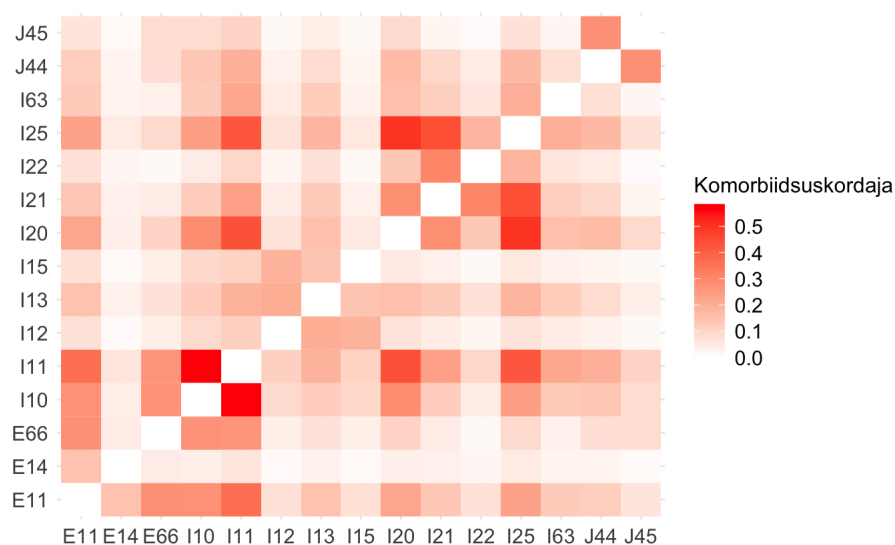
Niisiis jäi algsest 23-st haigusest peale filtreerimist alles 15 ning igasse andmetabelisse $C_{15}^2 = 105$ andmerida. Analüüsitavad haigused olid järgmised:

- E11 - insuliinisõltumatu suhkurtõbi,
- E14 - täpsustamata suhkurtõbi,
- E66 - rasvumus,
- I10 - hüpertooniatõbi ehk kõrgvererõhktõbi,

- I11 - südamekahjustusega hüpertooniatõbi,
- I12 - neerukahjustusega hüpertooniatõbi,
- I13 - südame- ja neerukahjustusega hüpertooniatõbi,
- I15 - sekundaarne hüpertensioon,
- I20 - stenokardia e rinnaangiin e rinnaahistus,
- I21 - äge müokardiinfarkt,
- I22 - korduv müokardiinfarkt,
- I25 - krooniline südame isheemiatõbi,
- I63 - peajuinfarkt,
- J44 - krooniline obstruktiivne kopsuhaigus,
- J45 - astma. (Sotsiaalministeerium, [kuupäev puudub](#))

3.2 Komorbiidsuse tabel kogu andmestiku peale

Lisaks komorbiidsuskordajate arvutamisele tegurite kaupa, leiti kordajad ka üldandmestikule valitud 15 haiguse puhul. Jooniselt 2 on näha, kuidas haiguse paar (I10, I11) on suurima kordajaga, mille väärtuseks on ligikaudselt 0,585. Kõrgete kordajatega (suuremad, kui 0,4) olid ka paarid (I20, I25), (I21, I25), (I20, I11) ning (I11, I25) vastavalt ligikaudsete väärtustega 0,502, 0,440, 0,434 ja 0,420. Niisiis olid väärtuse poolest esimesed 5 paari kõik kõrgvererõhkaigused ja südame isheemiatõved. Esimene paar, mis sisaldas ka mingit teist haigust oli (I11, E11) ligikaudse väärtusega 0,361 ehk lisaks kõrgvererõhkaigusele lisandus ka suhkrutõbi.



Joonis 2: Komorbiidsuskordajate väärtused üldandmestikus

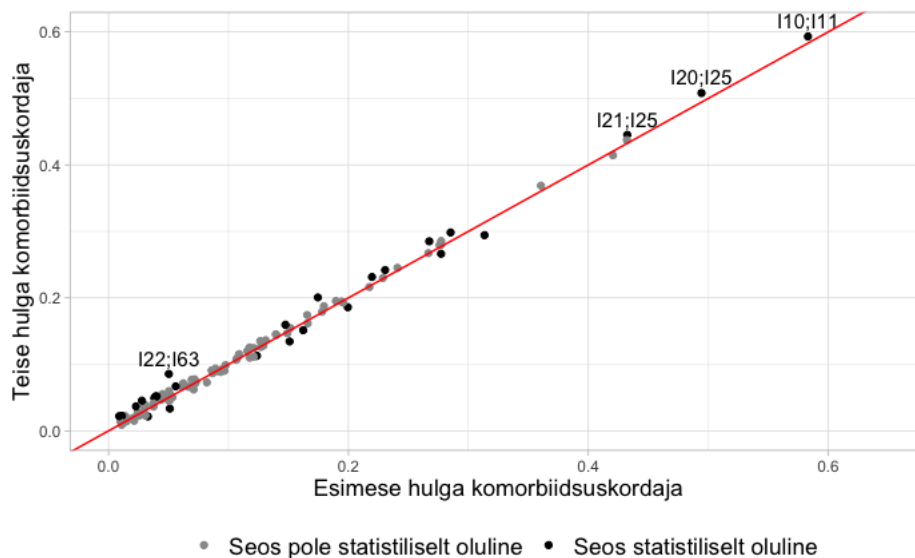
3.3 Komorbiidsuse tabelite võrdlus haiguste paaride kaupa

Komorbiidsuse tabelite võrdlemiseks ühendati tabelid vastavalt keskkonnategurile. Niisiis saadi 4 tabelit, mis sisaldasid iga haiguste paari kohta kahte komorbiidsuskordajat. Seejärel arvutati, kas kahe sõltumatu komorbiidsuskordaja vahe on statistiliselt oluline. Selleks rakendati Fisheri transformeeritud testi (Hinkle DE, 1988). Järgnevalt on keskkonnategurite kaupa toodud välja komorbiidsuskordajate vahelised seosed joonistena, kuhu on algselt sisse jäetud ka sellised haiguste paaride komorbiidsuskordajad, mille vaheline erinevus ei olnud statistiliselt olulised.

3.3.1 Kahe mittekattuva juhuslikult valitud hulga komorbiidsuskordajad

Komorbiidsuskordajate erinevused haiguste paaride kaupa olid juhuslikult valitud hulkade vahel nullilähedased. Jooniselt 3 on näha, et kõik väärtused on diagonaali

peal või selle lähedal. Statistiliselt mitteolulised väärtused olid enamjaolt juhtudel, kus mõlemad komorbiidsuskordajad olid alla 0,2. Kõige suurema kordajate erinevusega haiguste paar nende kahe hulga korral oli (I22,I63) ehk korduv infarkt ning peajuinfarkt, kus vahe oli ligikaudu 0,035. Kõikide kordajate erinevused jäid absoluutväärtuselt 0 – 0,035 vahele. Joonisel 3 on haiguste nimed lisatud nendele väärtustele, kus mõlemad komorbiidsuskordajad on suuremad kui 0,3 või komorbiidsuskordajate vahed on absoluutväärtuselt suuremad kui 0,03.



Joonis 3: Seos kahe mittekatuva juhuslikult valitud komorbiidsuskordajate vahel

3.3.2 Suitsetajate ja mittesuitsetajate hulkade komorbiidsuskordajad

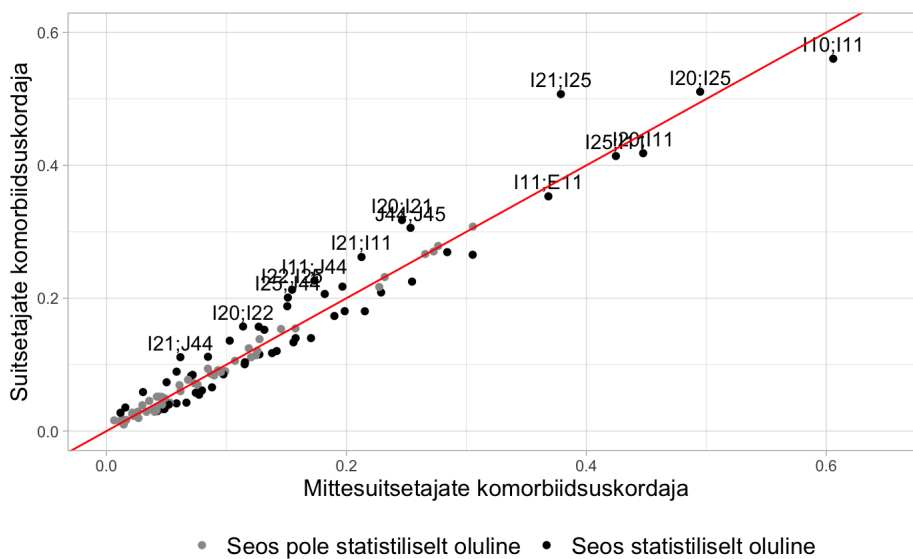
Suitsetamise puhul võrreldi komorbiidsuskordajaid mittesuitsetajate ja suitsetajate vahel. Suitsetajatena kaasati lisaks praegustele suitsetajatele ka need, kes olid varemalt suitsetajad. Peamiseks põhjuseks oli see, et käesolevasse analüüsi kaasati ka need haigused, mis diagnoositi enne Geenivaramuga liitumist ning Geenivaramul

ei ole teada, millal isik suitsetamisest loobus. Lisaks sellele võivad osad isikud suitsetamisest loobuda peale seda, kui neil mõni haigus diagnoositakse.

Kui juhuslikult valitud hulkade vahelised komorbiidsuskordajad haiguste paaride kaupa olid omavahel väikeste erinevustega ning väärtused paiknesid diagonaalil või selle läheduses, siis suitsetajate ja mittesuitsetajate vahelised statistiliselt olulised komorbiidsuskordajad haiguste paaride kaupa ei ole ühelgi juhul diagonaali peal. Esineb selliseid, mis asuvad diagonaali läheduses, kuid jooniselt 4 on näha, kuidas komorbiidsuskordajate erinevused on võrreldes joonisega 3 suuremad.

Joonisel 4 on haiguste nimed lisatud nendele väärtustele, kus mõlemad komorbiidsuskordajad on suuremad kui 0,3 või komorbiidsuskordajate vahed on absoluutväärtuselt suuremad kui 0,04. Kõik joonisel nimeliselt väljatoodud haiguste paaride väärtused, millel on erinevus suurem kui 0,04, asuvad ülalpool diagonaali. Ainus erand on haiguste paar (I10, I11) ehk kõrgvererõhktõbi ja südamekahjustusega kõrgvererõhktõbi allpool diagonaali, mille erinevus absoluutväärtuselt on 0,0453.

Suurima erinevusega on paar (I21, I25) ehk äge müokardiinfarkt ja krooniline südame isheemiatõbi, kus suitsetajate ja mittesuitsetajate komorbiidsuskordajad tulid ligikaudu vastavalt 0,51 ja 0,38 ning komorbiidsuskordajate erinevus ligikaudu 0,13.

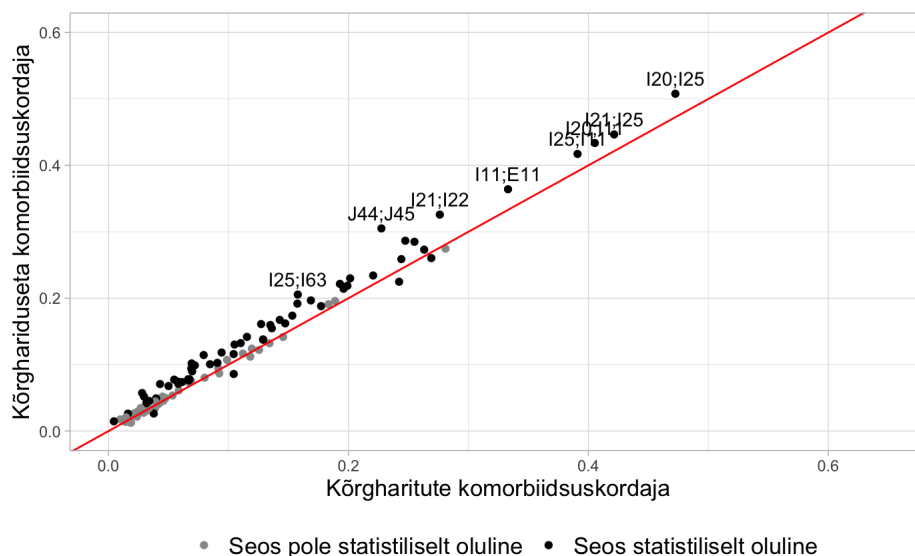


Joonis 4: Seos suitsetajate ja mittesuitsetajate komorbiidsuskordajate vahel

3.3.3 Kõrgharidusega ja kõrghariduseta hulkade komorbiidsuskordajad

Hariduse puhul võrreldi komorbiidsuskordajaid kõrgharidusega ja kõrghariduseta isikute vahel. Sarnaselt suitsetamisele on ka haridusel mõju komorbiidsuskordajatele. Jooniselt 5 on näha, kuidas enamus statistiliselt olulised seosed on ülalpool diagonaali. Allpool diagonaali olevatest haiguste paaridest on suurima komorbiidsuskordajate erinevusega paar (I10, I12) ehk kõrgvererõhktõbi ja neerukahjustusega kõrgvererõhktõbi, mille erinevus on absoluutväärtuselt ligikaudu 0,018.

Samamoodi nagu suitsetamise puhul, on ka hariduse juures haiguste nimed lisatud nendele väärtustele, kus mõlemad komorbiidsuskordajad on suuremad kui 0,3 või komorbiidsuskordajate vahed on absoluutväärtuselt suuremad kui 0,04. Kõige suurima komorbiidsuskordajate erinevusega oli paar (J44, J45) ehk krooniline obstruktiivne kopsuhaigus ja astma, kus kõrghariduseta ja kõrgharidusega komorbiidsuskordajad olid ligikaudu vastavalt 0,31 ja 0,23 ning nendevaheline erinevus absoluutväärtuselt ligikaudu 0,078.



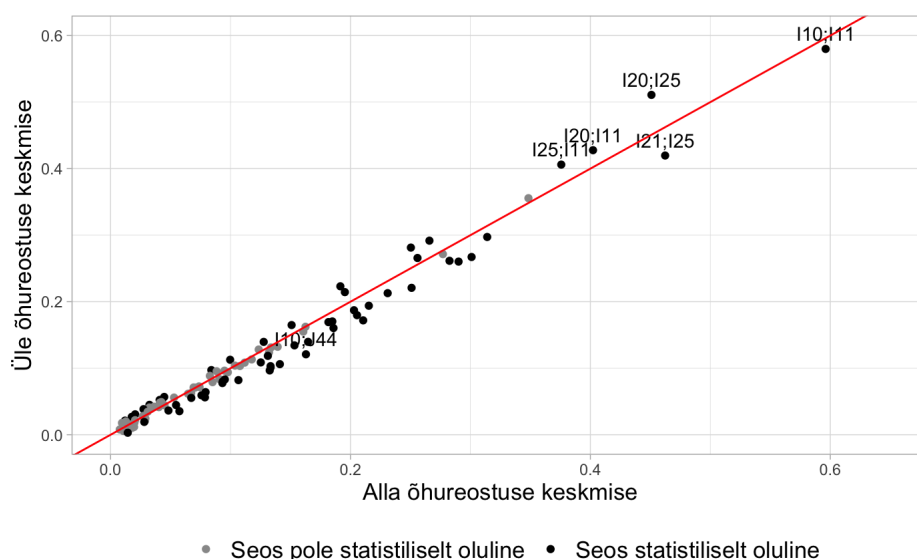
Joonis 5: Seos kõrgharidusega ja kõrghariduse ta hulkade komorbiidsuskordajate vahel

3.3.4 Ülipeente osakeste kontsentratsiooni järgi grupeeritud hulkade komorbiidsuskordajad

Õhureostuse (ülipeente osakeste $PM_{2,5}$ kontsentratsioon) puhul toimus grupeerimine komorbiidsustabelitesse vastavalt sellele, kas isiku kodukoha $PM_{2,5}$ kontsentratsioon oli alla või üle andmestiku keskmise $7,12 \mu g/m^3$. Kuigi algselt plaaniti andmed grupeerida vastavalt Maailma Terviseorganisatsiooni WHO 2021. aasta õhukvaliteedi juhendi tasemele $5 \mu g/m^3$ (World Health Organization, 2021), siis ainult 394 isiku puhul oli $PM_{2,5}$ kontsentratsioon alla või võrdne WHO fikseeritud tasemega. Seetõttu otsustati grupeerida vastavalt eelpool kirjeldatule.

Joonisel 6 on samuti haiguste nimed lisatud nendele väärtustele, kus mõlemad komorbiidsuskordajad on suuremad kui 0,3 või komorbiidsuskordajate vahed on absoluutväärtuselt suuremad kui 0,04. Jooniselt 6 on näha, et telgedel vahemikus 0 – 0,3 paiknevad paljud statistiliselt olulised väärtused allpool diagonaali, mille kaugus diagonaalist on samuti suurem, kui ülalpool diagonaali. See-eest suurima

komorbiidsuskordajate erinevusega oli paar (I20, I25) ehk stenokardia ja krooniline südame isheemiatõbi ligikauduselt absoluutväärtusega 0,060. Paarid (I21, I25) ehk äge müokardiinfarkt ja krooniline südame isheemiatõbi ning (I10, J44) ehk kõrgvererõhktõbi ja krooniline obstruktiivne kopsuhaigus olid suuruselt järgmiste erinevustega vastavalt 0,043 ja 0,042. Niisiis oli teguri $PM_{2,5}$ puhul rohkem neid haiguste paare, mille komorbiidsuskordaja oli suurem alla keskmise $PM_{2,5}$ väärtuse hulgas.



Joonis 6: Seos õhuresotuse kontsentratsiooni järgi grupeeritud hulkade komorbiidsuskordajate vahel

3.4 Komorbiidsusvõrgustikud

Komorbiidsusvõrgustike loomisesse kaasati iga teguri puhul ainult need haiguste paarid, mille komorbiidsuskordajate erinevus kahe hulga vahel oli statistiliselt oluline. Nii jäi alles teguri haridus puhul 62, teguri suitsetamise puhul 54, ülipeente osakeste kontsentratiooni $PM_{2,5}$ puhul 58 ja kahe juhuslikult valitud mittekattuva hulga puhul 26 haiguste paaride kombinatsiooni.

Analüüsi esimeses pooles komorbiidsuskordajaid võrreldes jagati andmed iga teguri puhul vastavalt fikseeritud tingimusele kaheks. Analüüsi teises, võrgustike loomise pooles, tekitati iga teguri kohta kaks võrgustikku vastavalt kahe hulga komorbiidsuskordajate erinevusele. Kordajatevaheline erinevus jagati kaheks olenevalt sellest, kas väärtus oli positiivne või negatiivne. Niisiis saadi igale tegurile kaks komorbiidsuskordajate erinevuste hulka, mille pealt võrgustikke luua.

Kõikide võrgustike puhul näitab serva paksus ja värvus kordajate erinevuse tugevust absoluutväärtuselt. Erinevuste absoluutväärtused jäid kõikide hulkade peale kokku vahemikku 0,07763 kuni 0,12856. Seega serva paksus ja värvus avaldub võrgustikel ühise suuruse ja värviskeemi järgi, kus kõige väiksemad väärtused saavad kollase, keskmised roheka ning kõige suuremad tumesinise värvuse (lisa 1 joonis 14). Värviskaala on lisatud juurde ka kõikide võrgustike joonistele. Serva paksust kujutakse graafil lineaarse seosega, kus miinimumväärtus 0,07763 saab ühe ühiku ning maksimumväärtus saab kümme ühikut serva paksuseks ning vahepealsed väärtused jaotuvad nende vahele (lisa 1 joonis 15).

Võrgustike loomiseks kasutati tarkvara Cytoscape (Shannon, Markiel, Ozier, Baliga, Wang, Ramage, Amin, Schwikowski ja Ideker, 2003) ning kõiki võrgustikke klasterdatakse algoritmiga GLay (Su, Kuchinsky, Morris, States ja Meng, 2010).

3.4.1 Suitsetajate ja mittesuitsetajate hulkade komorbiidsusvõrgustikud

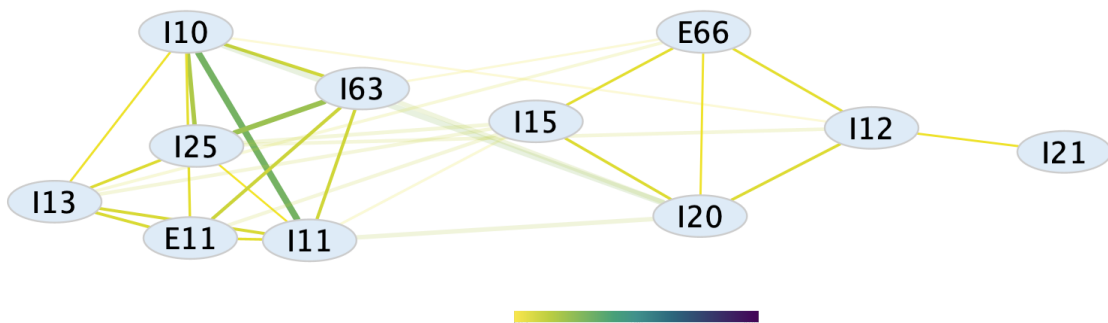
Suitsetamise puhul oli 30 haiguste paari komorbiidsuskordaja mittesuitsetajatel suurem ning 24 paaril oli suitsetajatel suurem. Mittesuitsetajate haiguste paarides oli 11 ning suitsetajatel 15 erinevat haigust.

Mittesuitsetajate oli suurima komorbiidsuskordajate erinevusega paar (I10,I11) väärtusega 0,0453 ning suuruselt järgmine oli (I20,I10) väärtusega 0,03984. Jooniselt 7 on näha mittesuitsetajate klasterdatud võrgustikku, kus GLay algoritmi rakendades on selgelt eristunud kaks klastrit. Esimeses vasakpoolses klastris on

järgmised 6 haigust: I10, I63, I25, I13, E11, I11. Teises klastris on haigused: E66, I15, I20, I12, I21. Suurima astmete arvuga ehk väärtusega 7 on haigused I10 ja I11. Kõige väiksema klastrite arvuga on I21, mis on seotud ainult haigusega I12.

Esimeses klastris on kõrgvererõhkaigustest kõrgvererõhktõbi, lisaks ka südamekahjustusega kõrgvererõhktõbi kui ka südame- ja neerukahjustusega kõrgvererõhktõbi. Samuti ka krooniline südame isheemiatõbi, peajuinfarkt ning insuliinisõltumatu suhkrutõbi.

Teises klastris on kõrgvererõhkaigustest neerukahjustusega kõrgvererõhktõbi ning sekundaarne hüpertensioon. Südame isheemiatõvedest stenokardia ning äge müokardiinfarkt. Haigustest on lisaks veel rasvumus.



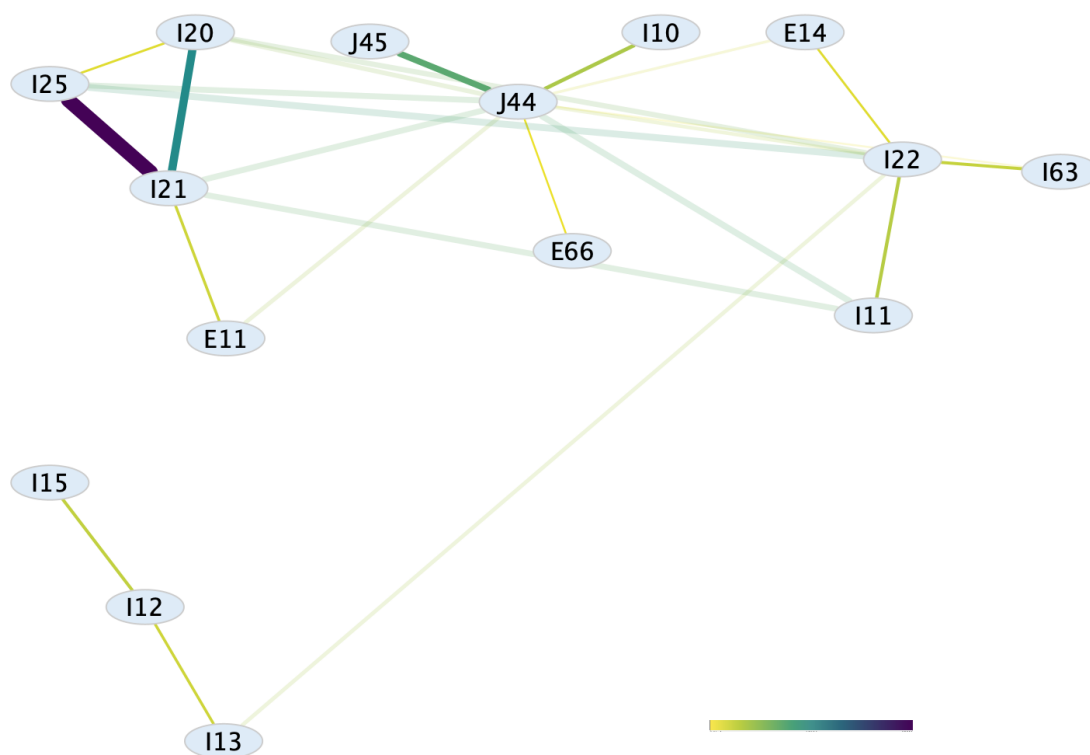
Joonis 7: Erinevuste võrgustik, kus mittesuitsetajate kordaja on suurem

Suitsetajatel oli suurima komorbiidsuskordajate erinevusega paar (I21, I25) väärtusega 0,12856 ehk kõikide tegurite hulkade suurima erinevusega. Jooniselt 8 on näha suitsetajate klasterdatud võrgustikku, kus GLay algoritmi rakendades on saadud 4 erinevat klastrit.

Kõige suurema astme arvuga on J44 väärtusega 11. Lisaks on kõrge astme arvuga haigus I22 väärtusega 7. Nelja haiguse puhul oli astmete arv 1, milleks olid E66, I10, J45, I15. Nendest esimesed kolm olid kõik ühenduses ainult haigusega J44.

Esimesse klastrisse kuuluvad haigused I11, I22, I63 ja E14 ehk südamekahjustu-

sega kõrgvererõhktõbi, korduv müokardiinfarkt, peajuinfarkt ning täpsustamata suhkrutõbi. Teises klastris on haigused J44, J45, I10 ja E66 ehk krooniline obstruktiivne kopsuhaigus, astma, kõrgvererõhktõbi ja rasvumus. Kolmandas klastris on haigused I20, I21, I25 ja E11 ehk stenokardia, äge müokardiinfarkt, krooniline südame isheemiatõbi ning insuliinisõltumatu suhkrutõbi. Neljandas, pisut eraldatud klastris on haigused I12, I13 ja I15 ehk neerukahjustusega kõrgvererõhktõbi, südame- ja neerukahjustusega kõrgvererõhktõbi ning sekundaarne hüpertensioon. Seega on viimases klastris ainult kõrgvererõhkaigused.



Joonis 8: Erinevuste võrgustik, kus suitsetajate kordaja on suurem

Kõik võrgustikus 7 esinevad haigused leiduvad ka võrgustikus 8. Võrgustikus 8 on lisaks veel haigused J44, J45, E14, I22 ehk mõlemad analüüsi kaasatud alumiste hingamisteede kroonilised haigused, täpsustamata suhkrutõbi ning korduv müo-

kardiinfarkt. Lisaks on J44 ehk krooniline obstruktiivne kopsuhaigus võrgustikus 8 kõige suuremate astmete arvuga, millest võime järeldada, et suitsetajatel on krooniline obstruktiivne kopsuhaigus paljude haiguste puhul kõrgem.

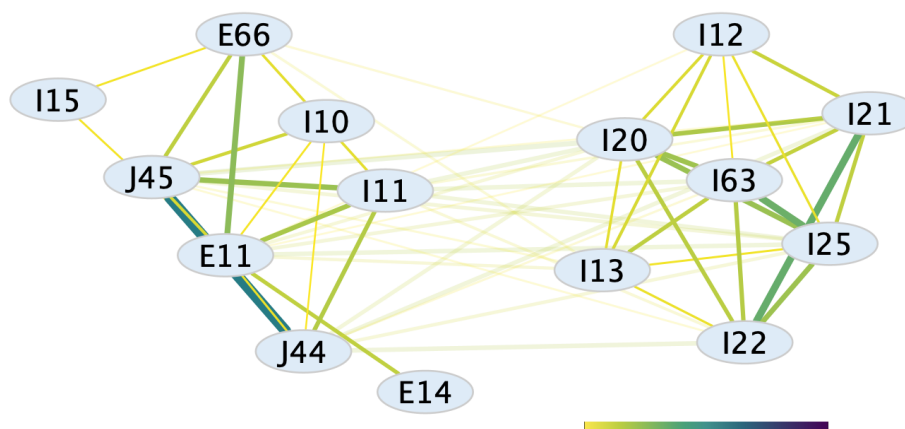
3.4.2 Kõrgharidusega ja kõrghariduseta hulkade komorbiidsusvõrgustikud

Hariduse puhul oli 57 haiguste paari komorbiidsuskordaja kõrghariduseta isikute hulgal suurem, 5 paaril kõrgharidusega isikute hulgal suurem. Kõrghariduseta hulga haiguste paarides oli 15 ning kõrgharidusega hulgal 7 erinevat haigust.

Jooniselt 9 on näha kõrghariduseta komorbiidsuskordajate erinevuste võrgustikku, kus GLay algoritmi rakendades on eristunud kaks klastrit. Suurima kordaja erinevuste väärtusega on paar (J44, J45) väärtusega 0,07763 ning suuruselt järgmine on paar (I21, I22) väärtusega 0,04968. Haigus E11 on suurima astme arvuga, milleks on 13. Suured astmete arvud on ka haigustel I20, J45, I25 vastavalt väärtustega 11, 10, 10. Kõige väiksema astme arvuga on haigus E14 väärtusega 1.

Esimeses klastris on järgmised 8 haigust: J45, J44, I10, I11, I15, E11, E14, E66 ehk alumiste hingamisteede kroonilised haigused, kõrgvererõhktõbi ja südamekahjustusega kõrgvererõhktõbi, sekundaarne hüpertensioon, nii insuliinisõltumatu kui ka täpsustamata suhkrutõbi ning rasvumus.

Teises klastris on järgmised 7 haigust: I12, I13, I20, I21, I22, I25, I63 ehk neerukahjustusega kõrgvererõhktõbi, südame- ja neerukahjustusega kõrgvererõhktõbi, kõik analüüsi kaasatud südame isheemiatõved ning peaajuinfarkt.



Joonis 9: Erinevuste võrgustik, kus kõrghariduseta kordaja on suurem

Kõrgharidusega komorbiidsuskordajate erinevuste võrgustiku 5 haiguste paari puhul oli haiguste paaride arv liiga väike, et GLay algoritmi oleks klasterdamiseks vaja kasutada. Erinevuste väärtused jäid vahemikku 0,00867 – 0,01848 ehk ükski haiguste paar ei olnud komorbiidsuskordaja poolest märgatavalt suurem kui kõrghariduseta komorbiidsuskordaja puhul. Suurima astmete arvuga oli haigus I10 ehk kõrgvererõhktõbi väärtusega 4, mis oli 7 haiguse peale ainuke suurema astmete arvuga kui 1. Võrgustik on väljatoodud lisas 1.

3.4.3 Ülipeente osakeste kontsentratiooni järgi grupeeritud hulkade komorbiidsusvõrgustikud

Teguri ülipeente osakeste kontsentratsioon $PM_{2,5}$ puhul oli 37 haiguste paari komorbiidsuskordaja üle keskmise ülipeente osakeste kontsentratiooni $PM_{2,5}$ hulgal suurem, 21 paaril oli alla keskmise ülipeente osakeste kontsentratiooni $PM_{2,5}$ hulgal suurem. Erinevaid haiguseid haiguste paaride hulkades oli vastavalt 14 ning 12.

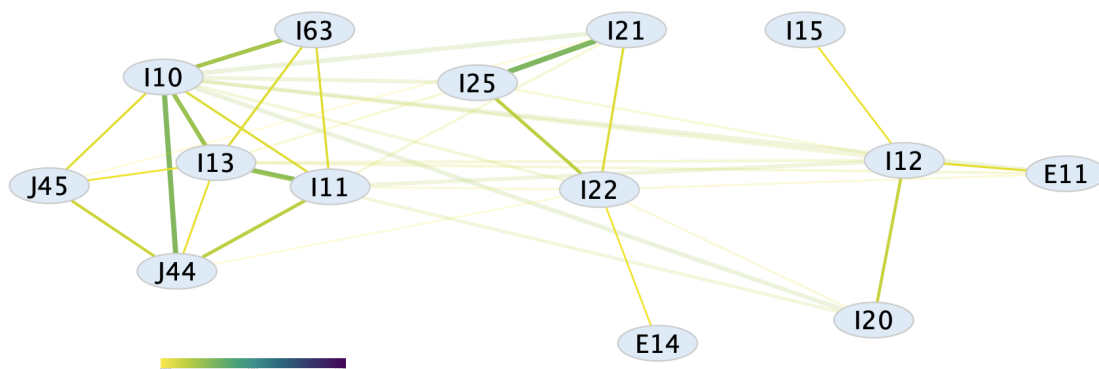
Joonisel 10 on alla keskmise $PM_{2,5}$ erinevuste võrgustikus tekkis GLay algoritmi rakendades 3 klastrit. Suurimate astemete arvudega on haigused I10, I13, I22, I12, I11 väärtustega 11, 9, 8, 7, 7. Kõige väiksema astme arvuga olid haigused E14 ja I15

väärtusega 1. Suurima erinevuse väärtusega 0,04265 on paar (I21, I25). Suuruselt järgmine on paar (I10, J44) väärtusega 0,04192.

Esimesse vasakpoolsesse klastrisse on haigused J44, J45, I10, I11, I13, I63 ehk alumiste hingamisteede kroonilised haigused, kõrgvererõhktõbi, lisaks ka südamekahjustusega, aga ka südame- kui ka neerukahjustusega kõrgvererõhktõbi ning peaajuinfarkt.

Järgmisesse klastrisse kuuluvad I21, I22, I25, E14 ehk äge müokardiinfarkt, korduv müokardiinfarkt, krooniline südame isheemiatõbi ning täpsustamata suhkrutõbi.

Viimases klastris on I12, I15, I20, E11 ehk neerukahjustusega hüpertooniatõbi, sekundaarne hüpertensioon, stenokardia ning insuliinisõltumatu suhkrutõbi.



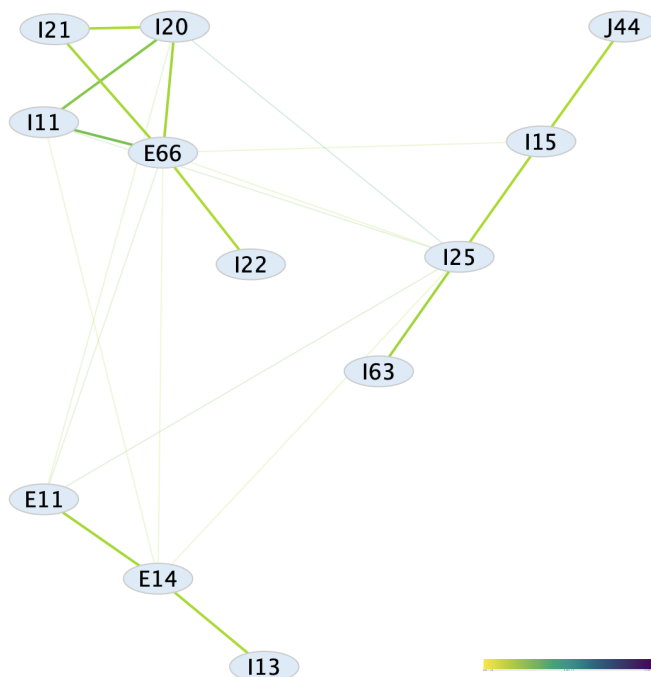
Joonis 10: Erinevuste võrgustik, kus alla $PM_{2,5}$ keskmise kordaja on suurem

Üle keskmise $PM_{2,5}$ erinevuste võrgustiku puhul on jooniselt 11 näha GLay rakendamisel kolme klastrit, mis on võrreldes joonisega 10 üksteisest eraldunumad. Suurima astmete arvuga on E66 astme arvuga 8. Haigused I25, I20, E14 on vastavalt väärtustega 7, 5, 5. Astme arvuga 1 on I63, I13, J44, I22. Suurima erinevuse on paar (I20, I25) väärtusega 0,0599.

Esimesse klastrisse kuuluvad I11, I20, I21, I22, E66 ehk südamekahjustusega kõrgvererõhktõbi, stenokardia, äge ja korduv müokardiinfarkt ning rasvumus.

Järgmisesse klastrisse kuuluvad I15, I25, I63, J44 ehk sekundaarne hüpertensioon, krooniline südame isheemiatõbi, peajuinfarkt ning krooniline obstruktiivne kopsuhaigus.

Viimasesse klastrisse kuuluvad I13, E11, E14 ehk südame- ja neerukahjustusega kõrgvererõhktõbi, insuliinisõltumatu ning täpsustamata suhkurtõbi.



Joonis 11: Erinevuste võrgustik, kus üle $PM_{2,5}$ keskmise kordaja on suurem

Haiguste, nagu näiteks astma, suurem komorbiidsuskordaja alla $PM_{2,5}$ keskmise hulgas paljude haiguste korral, viitab sellele, et $PM_{2,5}$ võib olla seotud teiste faktoritega, mis mängivad lisaks ülipeente osakeste kontsentratsioonile olulist rolli. Seetõttu tuleks vaadata näiteks mitme eksposoomitunnuse kombinatsiooni mõju. Näiteks on suurem õhureostus linnades, kuid samal ajal võivad olla linnas tagatud paremad tervise- ja haridusvõimalused. Samuti võis olla olukord, kus isik kolis peale diagnoosi saamist parema õhukvaliteediga asukohta, sest hetkel on teadmata,

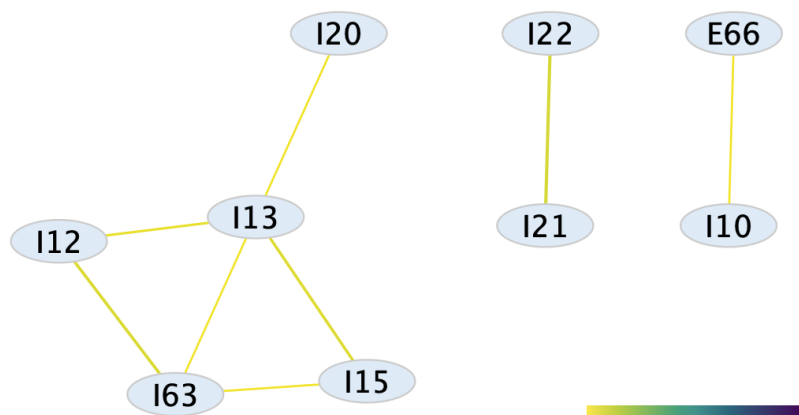
kas isik diagnoositi enne või pärast elukoha vahetamist.

3.4.4 Kahe mittekattuva juhuslikult valitud hulga komorbiidsusvõrgustikud

Kahe juhuslikult valitud mittekattuva hulga puhul oli 8 paari korral esimese hulga komorbiidsuskordaja suurem ning 18 paari puhul teise hulga komorbiidsuskordaja suurem. Esimese olukorra puhul oli 8 paari puhul 9 erinevat haigust ning teise olukorra puhul oli 18 paari seas 14 erinevat haigust.

Jooniselt 12 näha, et esimesse võrgustikku kaasatud 8 haiguste paari puhul ei muutunud GLay rakendamine haiguste paiknemist võrgustikus. Lisaks jäid erinevuste väärtused vahemikku 0,01085 – 0,01917 ehk väärtused ei olnud suured ning nende varieeruvus oli väike. Kahe paari (I21, I22) ehk ägeda müokardiinfarkti ja korduva müokardiinfarkti ning (I10, E66) ehk kõrgvererõhktõve ja rasvumuse puhul leidsid võrgustikus komorbiidsuskordajate vahed ainult paaridel endal.

Ainuke suurem klaster sisaldas haiguseid I12, I13, I15, I20, I63 ehk neerukahjustusega hüpertooniatõbi, südame- ja neerukahjustusega hüpertooniatõbi, sekundaarne hüpertensioon, stenokardia ja peaaajuinfarkt. Võrgustiku suurima astme arvuga haigus I13 paiknes selles klastris astme arvuga 4.



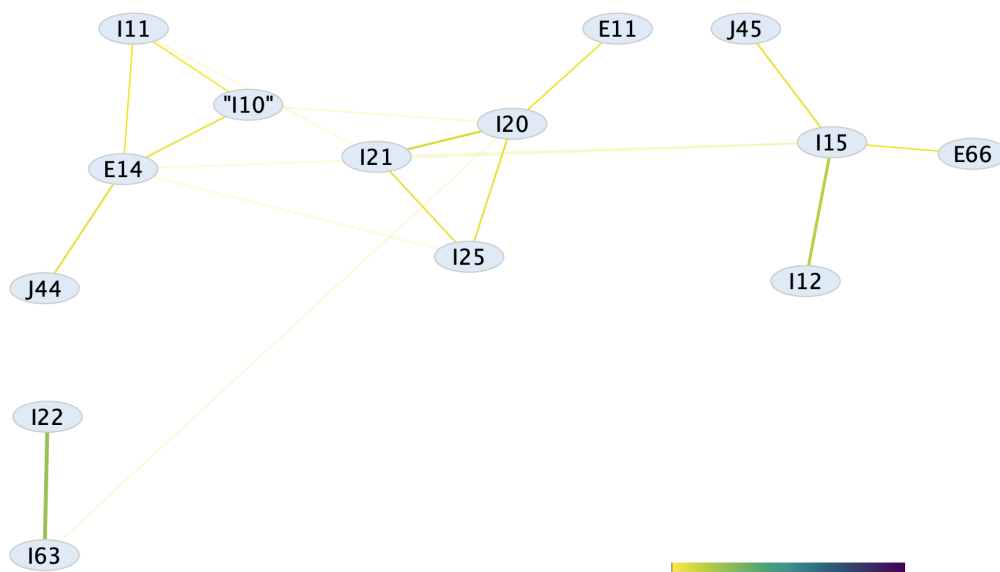
Joonis 12: Erinevuste võrgustik, kus esimese mittekattuva hulga kordaja on suurem

Teises võrgustikus tekkis klastreid rohkem. Jooniselt 13 on näha kolme suuremat ja ühte kahe haigusega klastrit. Kõige väiksemas klastris oli paar (I22, I63) ehk korduv müokardiinfarkt ja peajuinfarkt. Suurimad erinevused olid paaridel (I22, I63) ja (I12, I15) vastavalt 0,03537 ja 0,02615. Ülejäänud erinevused olid vahemikus 0,00999 – 0,01768. Üleüldiselt oli varieeruvus suurem kui võrgustiku 12 korral. Suurima astmete väärtusega 5 olid haigused E14, I15, I20.

Kolmest suuremast klastrist esimeses olid haigused I10, I11, J44, E14 ehk hüpertooniatõbi, südamekahjustusega hüpertooniatõbi, krooniline obstruktiivne kopsuhaigus ja täpsustamata suhkrutõbi.

Järgmises klastris olid I20, I21, I25, E11 ehk stenokardia, äge müokardiinfarkt, krooniline südame isheemiatõbi ning insuliinisõltumatu suhkrutõbi.

Viimasesse klastrisse kuulusid I12, I15, J45, E66 ehk neerukahjustusega hüpertooniatõbi, sekundaarne hüpertensioon, astma ja rasvumus.



Joonis 13: Erinevuste võrgustik, kus teise mittekattuva hulga kordaja on suurem

3.5 Komorbiidsuskordajate erinevuste võrdlus tegurite kaupa

Järgneva alapeatüki joonised on väljatoodud lisa 2. Joonisel 19 on näidatud komorbiidsuskordajate erinevused haiguste paaride kaupa vastavalt igale tegurile. Joonise loomiseks on kasutatud paketti ComplexHeatmap, mis vaikimisi kasutab dendrogrammi loomiseks hierarhilist klasterdamist kaugusmõõduga eukleidiline kaugus (Gu, 2022).

Haiguste paaride komorbiidsuskordajate erinevuste omavahelisel võrdlemisel tegurite kaupa on jooniselt 19 näha, kuidas kõikidele statistiliselt mitteolulistele erinevustele on fikseeritud väärtuseks 0. Samuti esineb palju selliseid haiguste paare, kus erinevus on oluline ainult ühe kindla tunnuse puhul.

Tunnuste suitsetamine ja haridus puhul tuleb jooniselt 19 hästi esile ülemiste haiguste paaride puhul, kuidas väärtused on vastandlikud, sest suitsetamise puhul leiti

erinevus lahutades suitsetajate kordajalt mittesuitsetajate kordaja ning hariduse puhul kõrgharidusega kordajalt kõrghariduseta kordaja. Suur osa ülemistest paaridest kuuluvad klastrisse tähisega A, mida sisaldavate paaride haigused on välja toodud joonisel 17. Klastrist A ei kaasatud võrgustikku haiguste paarid (E14, J44) ning (I12, I15), sest nende erinevused ei olnud tunnuse haridus puhul statistiliselt olulised. Klastrisse kuuluvast 16 paarist 8 sisaldavad haigust J44 ehk kroonilist obstruktiivset kopsuhaigust ning 6 paaris on haigus I22 ehk korduv müokardiinfarkt, mis võib viidata sellele, et nende kahe haiguse puhul suitsetamine suurendab komorbiidsuskordajaid ülejäänute joonisel 17 esinevate haigustega ning kõrghariduse omamine vastupidiselt vähendab. Dendrogrammil 19 ülejäänud haiguste paaride puhul üldiselt analoogset erimärgilisust rohkem ei esinenud.

PM_{2,5} erinevuse leidmiseks lahutati üle PM_{2,5} kontsentratsiooni keskmise hulgast alla PM_{2,5} kontsentratsiooni keskmise hulk. Tunnuse suitsetamine ja PM_{2,5} puhul olid alumiste haiguste paaride puhul, millest suur osa kuulus klastrisse B, erinevused sarnased. Alumisse klastrisse kuulus 10 paari, millest võrgustikku 18 kaasati 8 paari, sest paarid (I11, I12) ja (I21, I10) ei olnud tunnuse suitsetamine puhul statistiliselt olulised. Jooniselt 18 on näha, et 8 haiguste paarist 7 esines I10 ehk kõrgvererõhktõbi. Niisiis võib see viidata, et nii suitsetamine kui ka keskmisest kõrgem PM_{2,5} kontsentratsioon suurendab I10 komorbiidsuskordajaid joonisel 18 olevate haigustega. Ülejäänud haiguste paaride puhul klastrid nende kahe tunnuse vahel nii sarnased (kui klatri B puhul) ei olnud.

ComplexHeatmapi poolt läbiviidud klasterdamine andis tulemuseks, et tunnused PM_{2,5} ja juhuslikult valitud hulkade erinevused on kõige sarnasemad, sest need tunnused klasterdatakse esimesena. Tunnus haridus lisatakse järgmisena klasterdades see nende kahega. Ning viimasena lisatakse tunnus suitsetamine, mis viitab sellele, et tunnus suitsetamine on ülejäänud kolmest tunnusest kõige erinevam. Tunnuse PM_{2,5} erinevuste sarnanemine enim juhuslikult valitud hulkade erinevustele viitab sellele, et tunnuse PM_{2,5} korral ei saa tuvastada märkimisväärset mõju komorbiidsusele.

Kokkuvõte

Bakalaureusetöö eesmärk oli uurida keskkonnategurite mõju komorbiidsuse võrgustikele, vaadates juba varem uuritud kahte tegurit ning lisades ühe uudse eksposoomi tunnuse EXPANSE projektist. Uuritavateks teguriteks olid suitsetamine, haridus ning peenosakeste kontsentratsioon $PM_{2,5}$. Komorbiidsuse võrgustikud arvutati Eesti geenivaramu andmete põhjal.

Vaadeldavad 14 haigust olid kardiometaboolsed ja kopsuhaigused, mille põhjal loodi iga teguri kohta kaks komorbiidsustabelit, mis sisaldasid komorbiidsuskordajaid haiguste paaride kaupa. Iga teguri puhul võrreldi saadud komorbiidsuskordajate vahelist seost. Lisaks leiti ka komorbiidsuskordajad andmestikule, mille põhjal saadi, et haiguste paar (I10, I11) on suurima kordajaga väärtusega 0,585 ning üleüldiselt suurimad komorbiidsuskordajad olid kõrgvererõhkhaigustel ja südame isheemiatõvedel.

Teguri suitsetamine puhul tuli selgelt esile, kuidas suuremad kordajad kippusid olema suitsetajate hulgal. Suurima erinevusega oli paar (I21, I25) ehk äge müokardiinfarkt ja krooniline südame isheemiatõbi, kus erinevus tuli kõikide tegurite peale suurim ligikaudu väärtusega 0,13.

Teguri haridus puhul tuli hulkadevahelises komorbiidsuskordajate võrdluses erinevus isegi selgemalt esile kui suitsetamise puhul. Nimelt ainult 5 paari puhul 62-st tuli kordaja kõrgharidusega isikute hulgal suurem.

Peenosakeste $PM_{2,5}$ kontsentratsiooni puhul ei tulnud välja selget erinevust kahe hulga vahel. Kordajate vahelised erinevused olid omavahel üpris sarnased ning oli rohkem neid haiguste paare, mille kordaja oli suurem hulgas, kus $PM_{2,5}$ kontsentratsioon oli alla keskmise.

Võrgustike loomisel tuli teguri suitsetamine puhul erinevuste võrgustikus, kus suitsetajate kordaja oli suurem, esile haigus J44 ehk krooniline obstruktiivne kopsuhaigus, mis oli ühenduses 11 haigusega. Lisaks sisaldas erinevuste võrgustikus, kus

suitsetajate kordaja oli suurem, ka kõiki neid haiguseid, mis sisaldasid võrgustikus, kus mittesuitsetajate kordaja oli suurem.

Hariduse puhul oli peamiseks analüüsitavaks võrgustikuks see, mille puhul oli kõrghariduseta isikute hulga kordaja suurem. Võrgustikus oli suurima astmete arvuga E11 ehk insuliinisõltumatu suhkurtõbi väärtusega 13. Haiguste paar (J44, J45) ehk alumiste hingamisteede kroonilised haigused olid võrgustiku suurima erinevusega ligikaudse väärtusega 0,08.

Peenosakeste $PM_{2,5}$ kontsentratsiooni korral kahe võrgustiku loomisel oli E66 ehk rasvumus ainuke haigus, mis esines võrgustikus, kus üle $PM_{2,5}$ keskmise kordaja oli suurem, aga teises võrgustikus mitte. Võrgustikus, kus alla $PM_{2,5}$ keskmise kordaja oli suurem, esinesid võrreldes esimese võrgustikuga lisaks haigused nagu J45, I10 ning I12 ehk astma, kõrgvererõhktõbi ning neerukahjustusega kõrgvererõhktõbi. Tulemus, et astma on paljude haiguste korral suurema komorbiidsuskordajaga alla $PM_{2,5}$ keskmise hulgas, viitab sellele, et $PM_{2,5}$ võib olla seotud teiste faktoritega, mis mängivad lisaks olulist rolli ning vaadata tuleks näiteks mitme eksposoomitunnuse kombinatsiooni mõju. Samuti on juba varem näidatud astma ja õhureostuse seoseid ning astmahood võivad sageda halva õhukvaliteedi korral, seega on võimalik, et astmaatikud on juba enne Geenivaramuga liitumist kolinud võimalusel puhtama õhuga piirkondadesse.

Samamoodi dendrogrammil erinevuste võrdlemisel tegurite kaupa saadi tulemuseks, et klasterdamisel tunnused $PM_{2,5}$ ja juhuslikult valitud hulkade erinevused on kõige sarnasemad, mis viitab sellele, et $PM_{2,5}$ puhul ei tuvastatud märkimisväärset mõju komorbiidsusele.

Rakendades töös kirjeldatud metoodikat saab tulevikus arvutada komorbiidsuskordajate erinevused kõikide haiguste puhul erinevate eksposoomide kaupa ning samuti ka mitme eksposoomi kombinatsiooni jaoks. Seejärel koostada detailsemad mudelid tuvastatud haiguste klastrite jaoks, kasutades geneetilisi riskiskoore ja eksposoomi andmeid. Seda kõike saab eraldi teostada nii meeste kui ka naiste jaoks.

Kasutatud allikad

- Barabási, Albert-László ja Zoltán N. Oltvai (veebruár 2004). “Network biology: understanding the cell’s functional organization”. Nature Reviews Genetics 5.2, lk. 101–113. ISSN: 1471-0064. DOI: [10.1038/nrg1272](https://doi.org/10.1038/nrg1272).
- California Air Resources Board (2019). Inhalable Particulate Matter and Health (PM2.5 and PM10). (vaadatud 12.05.2024). URL: <https://ww2.arb.ca.gov/resources/inhalable-particulate-matter-and-health>.
- Corlin, Laura, Meghan I. Short, Ramachandran S. Vasani ja Vanessa Xanthakis (mai 2020). “Association of the Duration of Ideal Cardiovascular Health Through Adulthood With Cardiometabolic Outcomes and Mortality in the Framingham Offspring Study”. JAMA Cardiology 5.5, lk. 549. ISSN: 2380-6583. DOI: [10.1001/jamacardio.2020.0109](https://doi.org/10.1001/jamacardio.2020.0109).
- European Environment Agency (2022). Air quality in Europe 2022. Technical report 05/2022. European Environment Agency. DOI: [10.2800/488115](https://doi.org/10.2800/488115). URL: <https://www.eea.europa.eu/publications/air-quality-in-europe-2022>.
- Fan, Jingchun, Shulan Li, Chunling Fan, Zhenggang Bai ja Kehu Yang (september 2015). “The impact of PM2.5 on asthma emergency department visits: a systematic review and meta-analysis”. Environmental Science and Pollution Research 23.1, lk. 843–850. ISSN: 1614-7499. DOI: [10.1007/s11356-015-5321-x](https://doi.org/10.1007/s11356-015-5321-x).
- Goh, Kwang-Il, Michael E. Cusick, David Valle, Barton Childs, Marc Vidal ja Albert-László Barabási (mai 2007). “The human disease network”.

- Proceedings of the National Academy of Sciences 104.21, lk. 8685–8690.
ISSN: 1091-6490. DOI: [10.1073/pnas.0701361104](https://doi.org/10.1073/pnas.0701361104).
- Goryakin, Yevgeniy, Lorenzo Rocco ja Marc Suhrcke (august 2017). “The contribution of urbanization to non-communicable diseases: Evidence from 173 countries from 1980 to 2008”. Economics amp; Human Biology 26, lk. 151–163. ISSN: 1570-677X. DOI: [10.1016/j.ehb.2017.03.004](https://doi.org/10.1016/j.ehb.2017.03.004).
- Gu, Zuguang (august 2022). “Complex heatmap visualization”. iMeta 1.3. ISSN: 2770-596X. DOI: [10.1002/imt2.43](https://doi.org/10.1002/imt2.43).
- Hamra, Ghassan B., Francine Laden, Aaron J. Cohen, Ole Raaschou-Nielsen, Michael Brauer ja Dana Loomis (november 2015). “Lung Cancer and Exposure to Nitrogen Dioxide and Traffic: A Systematic Review and Meta-Analysis”. Environmental Health Perspectives 123.11, lk. 1107–1112. ISSN: 1552-9924. DOI: [10.1289/ehp.1408882](https://doi.org/10.1289/ehp.1408882).
- Hidalgo, César A., Nicholas Blumm, Albert-László Barabási ja Nicholas A. Christakis (aprill 2009). “A Dynamic Network Approach for the Study of Human Phenotypes”. PLoS Computational Biology 5.4. Toim. Lauren Ancel Meyers, e1000353. ISSN: 1553-7358. DOI: [10.1371/journal.pcbi.1000353](https://doi.org/10.1371/journal.pcbi.1000353).
- Hinkle DE Wiersma W, Jurs SG (1988). Applied statistics for the behavioral sciences. Boston: Houghton Mifflin Company.
- Hubálek, Zdenek (1982). “Coefficients of Association and Similarity, Based on Binary (Presence-Absence) Data: An Evaluation”. Biological reviews of the Cambridge Philosophical Society 4.57, lk. 671–672. URL: https://www.researchgate.net/publication/229695992_Coefficients_of_

[Association_and_Similarity_Based_on_Binary_Presence-Absence_Data_An_Evaluation.](#)

Koit, Mare (2022).

Graafiteooria Põhimõisted: Õppematerjal TÜ Teaduskooli õpilastele.
(vaadatud 29.03.2024). URL: https://teaduskool.ut.ee/sites/default/files/teaduskool/oppetoo/mat_gymn_graafiteooria.pdf.

Leitsalu, Liis, Toomas Haller, Tõnu Esko, Mari-Liis Tammesoo, Helene Alavere, Harold Snieder, Markus Perola, Pauline C Ng, Reedik Mägi, Lili Milani, Krista Fischer ja Andres Metspalu (veebuar 2014). “Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu”. International Journal of Epidemiology 44.4, lk. 1137–1147. ISSN: 1464-3685. DOI: [10.1093/ije/dyt268](https://doi.org/10.1093/ije/dyt268).

Morris, John H, Leonard Apeltsin, Aaron M Newman, Jan Baumbach, Tobias Wittkop, Gang Su, Gary D Bader ja Thomas E Ferrin (november 2011). “clusterMaker: a multi-algorithm clustering plugin for Cytoscape”. BMC Bioinformatics 12.1. ISSN: 1471-2105. DOI: [10.1186/1471-2105-12-436](https://doi.org/10.1186/1471-2105-12-436).

Murray, Christopher J L et al. (oktoober 2020). “Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019”. The Lancet 396.10258, lk. 1223–1249. ISSN: 0140-6736. DOI: [10.1016/s0140-6736\(20\)30752-2](https://doi.org/10.1016/s0140-6736(20)30752-2).

Münzel, Thomas, Mette Sørensen, Omar Hahad, Mark Nieuwenhuijsen ja Andreas Daiber (mai 2023). “The contribution of the exposome to the burden of cardiovascular disease”. Nature Reviews Cardiology 20.10, lk. 651–669. ISSN: 1759-5010. DOI: [10.1038/s41569-023-00873-3](https://doi.org/10.1038/s41569-023-00873-3).

National Institute for Occupational Safety and Health (2022).

Exposome and Exposomics. (vaadatud 12.05.2024). Centers for Disease

- Control ja Prevention. URL: https://archive.cdc.gov/www_cdc_gov/niosh/topics/exposome/default.html#:~:text=The%20exposome%20can%20be%20defined,%2C%20diet%2C%20lifestyle%2C%20etc.
- Orru, Hans, Erik Teinemaa, Taavi Lai, Tanel Tamm, Marko Kaasik, Veljo Kimmel, Kati Kangur, Eda Merisalu ja Bertil Forsberg (märts 2009). “Health impact assessment of particulate pollution in Tallinn using fine spatial resolution and modeling techniques”. Environmental Health 8.1. ISSN: 1476-069X. DOI: [10.1186/1476-069x-8-7](https://doi.org/10.1186/1476-069x-8-7).
- Patel, Chirag J. ja Arjun K. Manrai (2015). “Development of Exposome Correlations Globes to Map Out Environment-Wide Associations”. Teoses: Pac Symp Biocomput, 2015. Toim. Russ B. Altman, A. Keith Dunker, Lawrence Hunter, Teri E. Klein ja Marylyn D. Ritchie, lk. 231–242. URL: <http://psb.stanford.edu/psb-online/proceedings/psb15/patel.pdf>.
- Pavlopoulos, Georgios A, Maria Secrier, Charalampos N Moschopoulos, Theodoros G Soldatos, Sophia Kossida, Jan Aerts, Reinhard Schneider ja Pantelis G Bagos (aprill 2011). “Using graph theory to analyze biological networks”. BioData Mining 4.1. ISSN: 1756-0381. DOI: [10.1186/1756-0381-4-10](https://doi.org/10.1186/1756-0381-4-10).
- Rappaport, Stephen M. (aprill 2016). “Genetic Factors Are Not the Major Causes of Chronic Diseases”. PLOS ONE 11.4. Toim. Rodney John Scott, e0154387. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0154387](https://doi.org/10.1371/journal.pone.0154387).
- Sadegh, Sepideh, James Skelton, Elisa Anastasi, Andreas Maier, Klaudia Adamowicz, Anna Möller, Nils M. Kriege, Jaanika Kronberg, Toomas Haller, Tim Kacprowski, Anil Wipat, Jan Baumbach ja David B. Blumenthal (märts 2023). “Lacking mechanistic disease definitions and corresponding association data hamper progress in

- network medicine and beyond”. Nature Communications 14.1. ISSN: 2041-1723. DOI: [10.1038/s41467-023-37349-4](https://doi.org/10.1038/s41467-023-37349-4).
- Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski ja Trey Ideker (november 2003). “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks”. Genome Research 13.11, lk. 2498–2504. ISSN: 1088-9051. DOI: [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303).
- Sotsiaalministeerium (kuupäev puudub). RHK-10. (vaadatud 26.03.2024). URL: <https://rhk.sm.ee/>.
- Su, Gang, Allan Kuchinsky, John H. Morris, David J. States ja Fan Meng (november 2010). “GLay: community structure analysis of biological networks”. Bioinformatics 26.24, lk. 3135–3137. ISSN: 1367-4803. DOI: [10.1093/bioinformatics/btq596](https://doi.org/10.1093/bioinformatics/btq596).
- Tolonen, Hanna, Jaakko Reinikainen, Zhi Zhou, Tommi Härkönen, Satu Männistö, Pekka Jousilahti, Laura Paalanen, Annamari Lundqvist ja Tiina Laatikainen (juuli 2022). “Development of non-communicable disease risk factors in Finland: projections up to 2040”. Scandinavian Journal of Public Health 51.8, lk. 1231–1238. ISSN: 1651-1905. DOI: [10.1177/14034948221110025](https://doi.org/10.1177/14034948221110025).
- Tremblay, Johanne ja Pavel Hamet (november 2019). “Environmental and genetic contributions to diabetes”. Metabolism 100, lk. 153952. ISSN: 0026-0495. DOI: [10.1016/j.metabol.2019.153952](https://doi.org/10.1016/j.metabol.2019.153952).
- Valderas, J. M., B. Starfield, B. Sibbald, C. Salisbury ja M. Roland (juuli 2009). “Defining Comorbidity: Implications for Understanding Health and Health Services”. The Annals of Family Medicine 7.4, lk. 357–363. ISSN: 1544-1717. DOI: [10.1370/afm.983](https://doi.org/10.1370/afm.983).

- Vlaanderen, Jelle, Kees de Hoogh, Gerard Hoek, Annette Peters, Nicole Probst-Hensch, Augustin Scalbert, Erik Melén, Cathryn Tonne, G. Ardine de Wit, Marc Chadeau-Hyam, Klea Katsouyanni, Tõnu Esko, Karin R. Jongasma ja Roel Vermeulen (juuli 2021). “Developing the building blocks to elucidate the impact of the urban exposome on cardiometabolic-pulmonary disease: The EU EXPANSE project”. Environmental Epidemiology 5.4, e162. ISSN: 2474-7882. DOI: [10.1097/ee9.000000000000162](https://doi.org/10.1097/ee9.000000000000162).
- Voce, Antonio, Leach Anna, Wishart Ellen ja Duncan Pamela (2023). Air pollution in Europe. (vaadatud 12.05.2024). The Guardian. URL: <https://www.theguardian.com/environment/ng-interactive/2023/sep/20/europes-pollution-divide-see-how-your-area-compares>.
- Wang, Feng, Maggie Stewart, Sarah McDermott, Arminee Kazanjian, Bilkis Vissandjee, Marie DesMeules, Margaret de Groh ja Howard Morrison (jaanuar 2012). “Migration and Diabetes in British Columbia and Quebec: Prevalence and Health Service Utilization”. Canadian Journal of Public Health 103.1, lk. 59–64. ISSN: 1920-7476. DOI: [10.1007/bf03404070](https://doi.org/10.1007/bf03404070).
- Wheelock, Craig E. ja Stephen M. Rappaport (veebbruar 2020). “The role of gene–environment interactions in lung disease: the urgent need for the exposome”. European Respiratory Journal 55.2, lk. 1902064. ISSN: 1399-3003. DOI: [10.1183/13993003.02064-2019](https://doi.org/10.1183/13993003.02064-2019).
- World Health Organization (2020). Global health estimates: Leading causes of death. (vaadatud 08.05.2024). World Health Organization. URL:

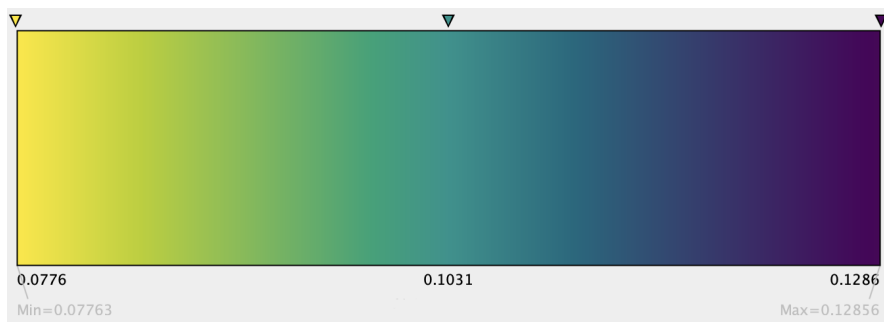
<https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-leading-causes-of-death>.

World Health Organization (2021).

What are the WHO Air quality guidelines? (vaadatud 12.05.2024). World Health Organization. URL: <https://www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines>.

– (2023). Noncommunicable diseases. (vaadatud 12.05.2024). World Health Organization. URL: <https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases>.

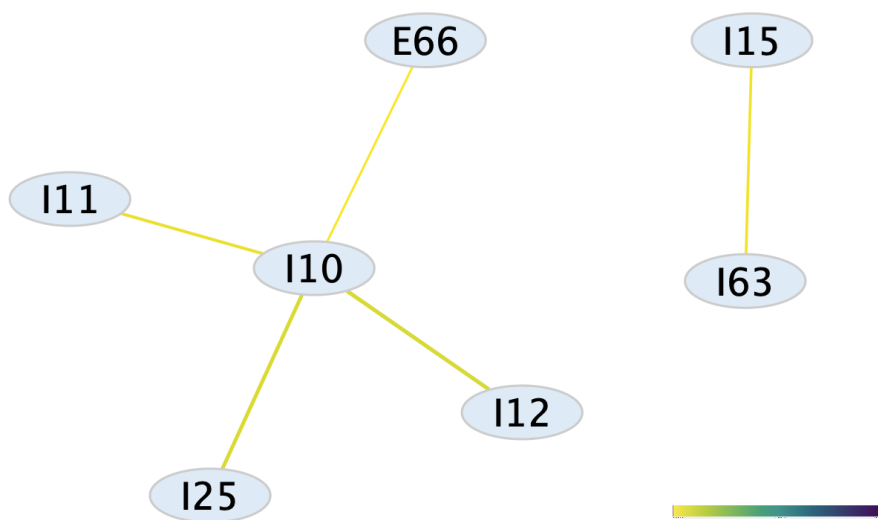
Lisa 1. Võrgustik



Joonis 14: Värviskeem võrgustiku servadele

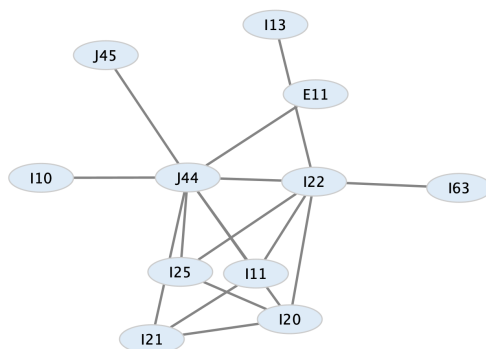


Joonis 15: Serva paksuse skeem võrgustiku servadele

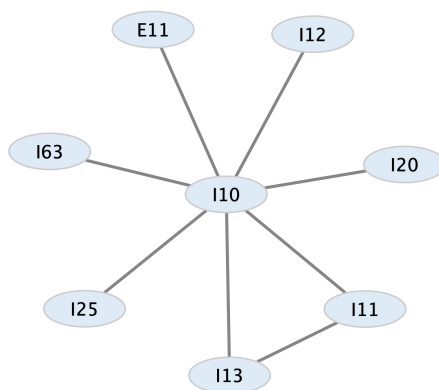


Joonis 16: Erinevuste võrgustik, kus kõrgharidusega kordaja on suurem

Lisa 2. Komorbiidsuskordajate erinevuste võrdlus



Joonis 17: Suitsetamise ja hariduse klastrisse A kuuluvad haiguste paarid



Joonis 18: Suitsetamise ja PM_{2,5} klastrisse B kuuluvad haiguste paarid

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Emma Räim,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose "Keskonnategurite mõju komorbiidsuse võrgustikele TÜ Eesti geenivaramu andmete põhjal", mille juhendajad on Jaanika Kronberg ja Krista Fischer, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 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.

Emma Räim

15.05.2024