

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Hanna Sõnajalg
**Inimestevaheliste kontaktide analüüs Eesti
CoMix uuringu põhjal**

Matemaatiline statistika
Bakalaureusetöö (9 EAP)

Juhendaja: Prof. Krista Fischer, Ph.D

TARTU 2022

INIMESTEVAHELISTE KONTAKTIDE ANALÜÜS EESTI COMIX UURINGU PÕHJAL

Bakalaureusetöö

Hanna Sõnajalg

Lühikokkuvõte

Koroonapandeemia tõi vajaduse uurida nakkushaiguste leviku seaduspärasusi ja nende põhjal planeerida nakkuse tõkestamise meetmeid. Et nakkus levib inimestevaheliste kontaktide kaudu, on oluline uurida, kui palju on inimestel erinevaid kontakte ja millest sõltub nende sagedus. Bakalaureusetöös analüüsitakse sotsiaalsete kontaktide uuringu CoMix andmeid, mis on kogutud Eestis 2021. aasta kevadel ja suvel. Töö eesmärk on leida mustreid eestlaste suhtlemises ja anda ülevaade, kuidas need uuringu käigus muutusid. Selleks luuakse vanusepõhised kontaktimaatriksid ja ehitatakse mudel, mille abil saab hinnata keskmist ööpäevaste kontaktide arvu.

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

Märksõnad: Kontaktimaatriksid, negatiivne binoomregressioon.

ANALYSIS OF SOCIAL CONTACTS IN ESTONIA FROM THE COMIX SURVEY

Bachelor thesis

Hanna Sõnajalg

Abstract

The coronavirus pandemic has highlighted the need to study the patterns of the spread of infectious diseases and to plan interventions according to them. Because the virus spreads by close person-to-person contact, it is important

to examine the frequency of different types of social contacts and what might affect it. The data being analyzed is collected from the social contacts survey CoMix in Estonia in the spring and summer of 2021. The aim of this Bachelor thesis is to find communication patterns of Estonians and to give an overview, how these patterns change during the study period. To this end, age-based contact matrices are created and a negative binomial regression model is built to estimate the average number of daily contacts.

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

Key Words: Contact matrices, negative binomial regression.

Sisukord

Sissejuhatus	4
1 Andmed	5
2 Statistiline metoodika	8
2.1 Kontaktimaatriksite konstrueerimine	8
2.2 Negatiivne binoomregressioon	9
3 Andmete analüüs	12
3.1 Osalejad	12
3.2 Kontaktide kirjeldus	16
3.3 Kontaktimaatriksid	22
3.4 Regressioonimudel päevas tehtud kontaktide arvule	26
Kokkuvõte	33
Kasutatud allikad	34

Sissejuhatus

Koroonaviirus levib inimeselt inimesele piisknakkuse kaudu. Nakatutakse, kui olakse lähedases kontaktis mõne nakatunud inimesega. Ainult selliste kontaktide piiramise muudab keeruliseks asjaolu, et nakatunud ei pruugi olla haigusele iseloomulikke sümptomeid ehk ta pole teadlik sellest, et ta on võimeline viirust edasi kandma. Seega tekib efektiivsete nakkuse tõkestamise meetmete rakendamiseks vajadus uurida, kui palju on inimestel erinevaid kontakte ja millest sõltub nende sagedus.

Suhtlemise seaduspärasusi saab hõlsalt visualiseerida kontaktimaatriksi abil. Inimesed on jagatud mingisuguse tunnuse põhjal n gruppi ja nende andmed kontaktide kohta on koondatud $n \times n$ ruutmaatriksisse, kus arv i . reas ja j . veerus tähistab keskmist kontaktide arvu i . ja j . grupi inimeste vahel.

CoMix on rahvusvaheline uuringukonsortsium, mille eesmärgiks on uurida sotsiaalsete kontaktide sagedust ja struktuuri erinevates Euroopa riikides. Bakalaureusetöö keskendub kontaktide uuringu Eesti andmete analüüsile.

Töö eesmärk on leida mustreid eestlaste suhtlemises ja anda ülevaade, kuidas need uuringu käigus muutusid. Selleks luuakse vanusepõhised kontaktimaatriksid ja ehitatakse mudel, mille abil saab hinnata keskmist ööpäevaste kontaktide arvu.

1 Andmed

Töös kasutatakse andmeid, mis on kogutud Tervise Arengu Instituudi poolt Eestis COVID-19 pandeemia uuringu käigus. Eesti uuring on osa suuremast programmist, mida rahastatakse Euroopa Liidu teadusuuringute ja innovatsiooni programmi „Horisont 2020“ projekti „EpiPose“ ja Ühendkuningriigi meditsiiniuuringute nõukogu (MRC) kaudu. Programmis osaleb peale Eesti veel 20 Euroopa riiki. Informatsioon uuringu korralduse kohta pärineb Tervise Arengu Instituudilt saadud küsitlusandmetest.

Uuring koosnes üheksast küsitluslainest: 7 täiskasvanute ja 2 alaealiste küsitluslainet. Uuringusse kutsututel paluti küsimustikule vastata ligikaudu kord iga kahe nädala jooksul. Osalemise võis katkestada igal ajal ja võimalik oli ka mõnele küsimustikule vastamine vahele jätta. Alla 18-aastased küsimustikule ise ei vastanud ning nende vastused märkis üles nende vanem või hooldaja.

Iga küsitluse korral pidid osalejad meelde tuletama kõik korrad, kui oldi kellegagi kontaktis vastamise eelnenud päeva hommikul kella viiest kuni vastamise päeva hommikul kella viieni. Kontakti all mõeldi ainult vahetuid kontakte. See tähendab, et teise inimesega kohtuti isiklikult ja vahetati vähemalt paar sõna või oldi füüsilises kontaktis (näiteks kätlemine, embamine, suudlemine või kontaktsordialad).

Iga kontakti puhul tuli märkida järgmised tunnused:

- kontakti kestus (vähem kui 5 min, 5-14 min, 15-59 min, 1-4 h, enam kui 4 h);
- kontakti tüüp - kodu, töö, kool, muu (vabaaja veetmise kohad, poed jne);
- kontaktisiku vanus (või ligikaudne vanusegrupp) ja sugu;
- kas kontaktisik kuulus vastaja leibkonda;
- kas kasutati maski, visiiri või muud viiruse eest kaitsvat vahendit.

Võimalik oli märkida kontakte ka agregeeritult. See tähendab, et kui vastajal polnud võimalik iga kontakti eraldi kirja panna, sest näiteks ta töökoht nõuab klientidega suhtlemist, siis sai ta märkida, ligikaudu mitme inimesega suhtles ta vanusegruppidest 0-18, 18-64 ja 65+. Jällegi tuli arvestada ainult neid inimesi, kellega vahetati vähemalt paar sõna.

Ülevaade küsitluslainetest on toodud tabelis 1.

Tabel 1: Ülevaade küsitluslainetest.

Küsitluslaine	Valim	Algus	Lõpp	Osalejaid	Kontakte
A1	Täiskasvanud	20. apr 2021	28. apr 2021	1547	9014
A2	Täiskasvanud	6. mai 2021	14. mai 2021	1314	8577
C1	Lapsed	8. juuni 2021	15. juuni 2021	297	4066
A3	Täiskasvanud	25. mai 2021	8. juuni 2021	1111	8900
A4	Täiskasvanud	23. juuni 2021	3. juuli 2021	947	6323
A5	Täiskasvanud	8. juuli 2021	22. juuli 2021	805	3938
A6	Täiskasvanud	26. juuli 2021	3. aug 2021	683	3915
A7	Täiskasvanud	10. aug 2021	17. aug 2021	579	2477
C2	Lapsed	10. aug 2021	23. aug 2021	299	1827

Uuringu alguses oli osalejate arv 1547, igas järgnevas täiskasvanute küsitluslaines oli osavõtjaid ligikaudu 200 võrra vähem ja seitsmendas laines oli vastanuid kolm korda vähem võrreldes uuringu algusega. Mõlemas alaealiste küsitluslaines oli osalejaid ligikaudu 300.

Uuringust osavõtjatelt küsiti nende vanust, sissetuleku suurusjärku, leibkonna suurst, haridustaset, samuti uuriti suhtumise kohta viiruse tõsidusse ja selle tõkesta-

miseks rakendatud meetmetesse.

Küsimuslainete puhul, kus valimis olid alla 18-aastased, paluti märkida, kas laps käis küsimustikule vastamisele eelnenud ehk kontaktide jälgimise päeval koolis või mitte. Kooli alla loeti kõik haridusasutused, kaasa arvatud lastesõimed, lasteaiad, kõrgkoolid.

Kogu uuringu jooksul märgiti kontakte kokku ligikaudu 50 000 ja suurima osalusega küsitluslaines A1 oli neid 9000 (vt Tabel 1). Kuigi kahes alaealiste küsitluslaines oli osalejate arv sama, siis augustis toimunud küsitluses oli osalejate kontaktide arv enam kui 2 korda väiksem võrreldes juunis toimunud küsitlusega.

2 Statistiline metoodika

Töös genereeritakse vanusepõhised kontaktimaatriksid. Samuti viiakse läbi regressioonanalüüs, mille eesmärk on vastanute tunnuste abil kirjeldada nende poolt ühes päevas tehtud kontaktide arvu.

2.1 Kontaktimaatriksite konstrueerimine

Sebastian Funk (2020) on kontaktimaatriksite tekitamiseks välja töötanud rakendustarkvara R paketi Socialmixr. Maatriksite loomiseks on vaja leida keskmine kontaktide arv kõigi vanusegruppide vahel. Keskmise kontaktide arvu varieeruvuse hindamiseks ehk näiteks sellele usaldusintervalli leidmiseks saab kasutada bootstrap-meetodit. Kuigi antud töös luuakse kontaktimaatriksid ilma keskmise kontaktide arvu varieeruvuse hindamiseta, siis tutvustatakse siiski, kuidas toimub bootstrap-meetodiga kontaktimaatriksite genereerimine.

Iga küsitluslaine korral võetakse juhuslikult tagasipanekuga valimeid selles küsitluslaines osalejate seast. Kui vastanu või kontaktisiku puhul puudub informatsioon vanuse kohta, jäetakse nad valimist välja. Võetud valimite arv on 50 ja ühe valimi suurus on võrdne konkreetsele küsimustikule vastanute arvuga. Seejärel valitakse juhuslikult tagasipanekuga kontaktid valimitesse sattunutele. Tulemuseks on maatriksid, mille ridades on osalejate ja veergudes kontaktisikute vanusegrupid. Vanusegruppe on kokku 9: 0-4, 5-11, 12-17, 18-29, 30-39, 40-49, 50-59, 60-69 ning 70-aastased või vanemad. Arv maatriksi i . reas ja j . veerus tähistab, palju kontakte oli osalejatel vanusegrupist i keskmiselt vanusegruppi j kuulunud isikutega. Seega genereeritakse iga küsimuslaine kohta 50 kontaktimaatriksit, mis lõpliku tulemuse saamiseks liidetakse kokku ja jagatakse arvuga 50.

Sellisel tekitatud vanusepõhised kontaktimaatriksid on lünklikud, sest täiskasvanute küsimuslainetes ei saa me valimisse osalejaid vanuses 0 kuni 18, ja küsitluslainetes, kus täiskasvanud on täitnud küsitluse oma laste kohta, ei satu valimisse

täiskasvanuid. Terviklike kontaktimaatriksite saamiseks kombineeritakse omavahel küsimuslaineid neljal erineval moel. Täiskasvanute andmeid kombineeritakse laste küsitluslainelega C1 ja C2 ning ka lastega, kes käisid või ei käinud küsitlusele eelneval päeval koolis. Kahe grupi andmete kombineerimine tähendab seda, et bootstrap-meetodil valimeid võttes võib sinna sattuda isikuid mõlemast grupist.

Andmete põhjal loodud maatriksid pole sümmeetrilised. See tähendab, et kontaktide arv, mis oli vanusegruppi 18-29 kuulujatel vanusegruppi 40-49 kuulujatega, ei pruugi tulla võrdne kontaktide arvuga, mis esines vanusegruppi 40-49 kuulujatel vanusegruppi 18-29 kuulujatega. Olgu i . reas ja j . veerus olev keskmine kontaktide arv c_{ij} ja tähistame inimeste koguarvu grupis i tähega N_i . Sümmeetrilise maatriksi korral kehtiks

$$N_i c_{ij} = N_j c_{ji}.$$

Selle saavutamiseks saab kasutada võrdust

$$c'_{ij} = \frac{1}{2N_i} (N_i c_{ij} + N_j c_{ji}),$$

kus c'_{ij} tähistab kontaktide arvu sümmeetrilises maatriksis.

Eesti rahvaarvud vanusegruppides on saadud Statistikaameti leheküljelt (ESA).

2.2 Negatiivne binoomregressioon

Negatiivse binoomregressiooni ülevaade põhineb Joseph M. Hilbe teosel (2011).

Kontaktide arv ööpäevas on diskreetne tunnus, mille võimalikud väärtused on mittenegatiivsed täisarvud. Kontaktide arv ehk uuritav tunnus tähistatakse sümboliga Y . Negatiivse binoomregressiooni mudeli korral eeldatakse, et Y on negatiivse binoomjaotusega ja et uuritava tunnuse keskvärtus avaldub funktsioonina argu-

menttunnustest:

$$\ln \mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p,$$

kus μ on uuritava tunnuse keskväärtus, β_0, \dots, β_p on hinnatavad parameetrid ja x_1, \dots, x_p on argumenttunnuste väärtused. Hinnatavate parameetrite vektor tähistatakse $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$.

Olgu valimimaht n . Vaadeldakse indiviidi i . Tema kohta on teada seletatavate tunnuste väärtused $x_{1i}, x_{2i}, \dots, x_{pi}$. Informatsioon indiviidide argumenttunnuste kohta koondatakse maatriksisse

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{p1} & x_{p2} & \cdots & x_{pp} \end{pmatrix}.$$

Maatriksi X i . rida tähistatakse sümboliga X_i ehk $X_i = (1, x_{i1}, \dots, x_{ip})$. Siis on logaritmi indiviidi i kontaktide arvu keskväärtusest võimalik maatrikskujul lahti kirjutada:

$$\ln \mu_i = X_i \beta,$$

millest

$$\mu_i = e^{X_i \beta}.$$

Funktsiooni, mis seab uuritava tunnuse Y väärtusele y_i vastavusse tema esinemise tõenäosuse $P(Y = y_i) = p(y_i)$, nimetatakse tõenäosusfunktsiooniks ja negatiivse

binoomjaotuse tõenäosusfunktsioon on esitatav kujul:

$$p(y_i) = P(Y = y_i) = \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha\mu_i} \right)^{1/\alpha} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i} \right)^{y_i} \quad i = 1, 2, \dots, n,$$

kus μ_i on tunnuse Y keskväertus i . indiviidi puhul ja $\alpha > 0$ on dispersiooni parameeter. Uuritava tunnuse Y dispersioon avaldatakse keskväertuse ja dispersiooni parameetri abil järgmiselt: $\mu + \alpha\mu^2$.

Seega on mudeli ehitamiseks vaja hinnata parameetrite α ja β väärtused. Seda tehakse kasutades suurima tõepära meetodit. Tõepärafunktsioon avaldub kujul:

$$L(\alpha, \beta) = \prod_{i=1}^n p(y_i) = \prod_{i=1}^n \frac{\Gamma(y_i + 1/\alpha)}{\Gamma(y_i + 1)\Gamma(1/\alpha)} \left(\frac{1}{1 + \alpha e^{X_i\beta}} \right)^{1/\alpha} \left(\frac{\alpha e^{X_i\beta}}{1 + \alpha e^{X_i\beta}} \right)^{y_i}.$$

Logaritmilise tõepärafunktsiooni kuju on

$$l(\alpha, \beta) = \sum_{i=1}^n \left(y_i \ln \alpha + y_i(X_i\beta) - \left(y_i + \frac{1}{\alpha} \right) \ln(1 + \alpha e^{X_i\beta}) + \right. \\ \left. + \ln \Gamma \left(y_i + \frac{1}{\alpha} \right) - \ln(y_i + 1) - \ln \Gamma \left(\frac{1}{\alpha} \right) \right).$$

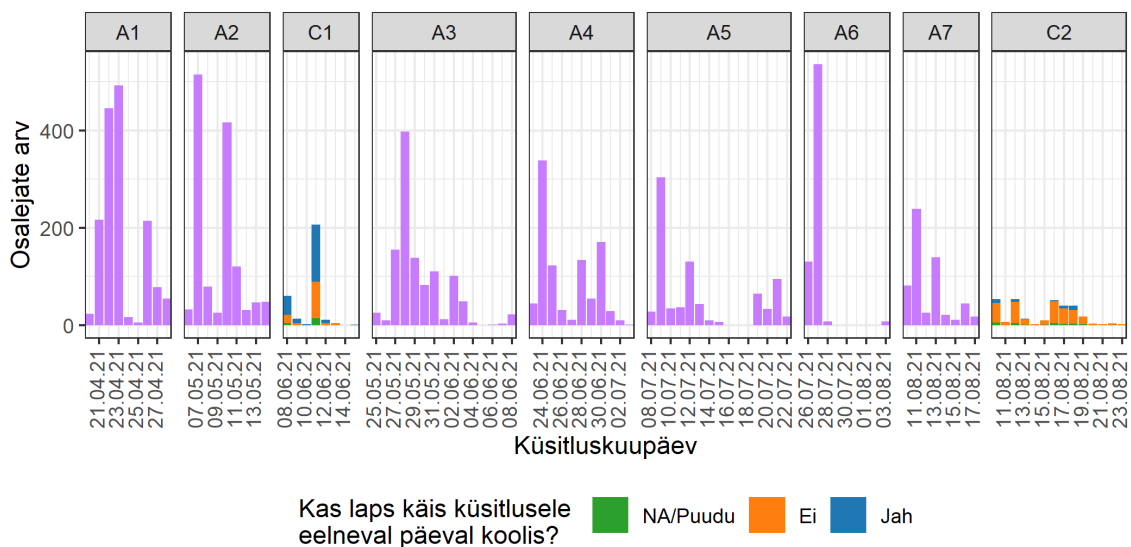
Maksimiseerides funktsiooni $l(\alpha, \beta)$ parameetrite α ja β järgi, leitaksegi neile suurima tõepära hinnangud.

3 Andmete analüüs

Alapeatükkides „Osalejad“, „Kontaktide kirjeldus“ ja „Kontaktimaatriksid“ toodud tabelid ja joonised on genereeritud keskkonnas GitHub publitseeritud koodi abil. Amy Gimma poolt avaldatud kood toetab CoMix uuringu andmetest ülevaate loomisel. (Gimma *et al.*, 2022)

3.1 Osalejad

Joonisel 1 on toodud osalejate arvud kuupäevade kaupa. Kuupäev iga osaleja puhul tähistab päeva, kui vastati küsimustikule, ehk üles märgitud kontaktid toimusid eelneval päeval. Laste küsitluslainetes on erinevate värvidega märgitud osalejad, kes käisid küsimustikule vastamiseks eelneval päeval haridusasutuses, ja kes mitte.



Joonis 1: Osalejate arv küsitluslainetes kuupäeviti.

Näha on, et kõigis küsitluslainetes on üks või kaks kõige suurema osalusega kuupäeva. Küsitluslaine A1 on need kuupäevad 22. ja 23. aprill, mis on vastavalt neljapäev ja reede. Laine A2 on 70% osalejatest vastanud küsitlusele 7. või 10. mail. Tegu on vastavalt reede ja esmaspäevaga. Nii lainetes A3 kui ka A5 oli suurima osalusega

reedesed päevad, laines A4 kolmapäev, laines A6 vastasid 80% osalejatest küsimustikule teisipäeval ning laines A7 vastasid enam kui 200 inimest küsimustikule kolmapäeval.

Alaealiste küsitluslaine C1 kõige populaarsem vastamise päev oli 11. juuni ehk reedene päev, teisel kohal oli 8. juuni ehk teisipäev. Nendel kahel päeval vastas küsimustikule 90% laine C1 osalejatest. Paljudes koolides õppetöö juunikuu esimeses pooles veel kestab ja 90% laine C1 osalejatest märkisid kontakte üles esmaspäeva ja neljapäeva kohta, mis on koolipäevad. Küsitluslaine C2 toimus augustis, kui paljud haridusasutused on suletud, seega on ka küsitlusele eelneval päeval kooliskäijaid väga vähe. Laines C2 oli küsimustikule vastajaid viiel päeval ligikaudu võrdne arv.

Teave osalejate vanuselise koosseisu kohta küsitluslaineti on toodud tabelis 2.

Tabel 2: Osalejate vanuseline jaotus lainetes.

Vanusegrupp	Küsitluslaine									
	A1	A2	A3	A4	A5	A6	A7	C1	C2	
0-4	-	-	-	-	-	-	-	50 (16,8%)	56 (18,7%)	
5-17	-	-	-	-	-	-	-	247 (83,2%)	234 (78,3%)	
18-29	242 (15,6%)	189 (14,4%)	166 (14,9%)	123 (13,0%)	48 (6,0%)	70 (10,2%)	81 (14,0%)	-	-	
30-39	270 (17,5%)	237 (18,0%)	197 (17,7%)	166 (17,5%)	81 (10,1%)	121 (17,7%)	113 (19,5%)	-	-	
40-49	255 (16,5%)	225 (17,1%)	182 (16,4%)	165 (17,4%)	123 (15,3%)	112 (16,4%)	98 (16,9%)	-	-	
50-59	286 (18,5%)	233 (17,7%)	200 (18,0%)	174 (18,4%)	171 (21,2%)	116 (17,0%)	101 (17,4%)	-	-	
60-69	271 (17,5%)	237 (18,0%)	201 (18,1%)	162 (17,1%)	193 (24,0%)	138 (20,2%)	99 (17,1%)	-	-	
70+	223 (14,4%)	193 (14,7%)	165 (14,9%)	157 (16,6%)	189 (23,5%)	126 (18,4%)	87 (15,0%)	-	-	
Puuduv	-	-	-	-	-	-	-	-	9 (3,0%)	

Üldiselt on kõigis täiskasvanute küsitluslainetes osalejate vanuseline jaotus ühtlane. Vastajate arv vanusegrupis moodustab peaaegu iga küsitluslaine korral kõigi osalejate arvust ligikaudu 16 kuni 17%. Veidi erilisem on laine A5, kus nooremaid osalejaid on vähem kui tavapäraselt. 18- kuni 39-aastased osalejad moodustavad kõigist osalejatest vaid 16%.

Mõlemas alaealiste küsitluslaines on 5- kuni 17-aastaseid osalejaid ligikaudu 80% kõigist osalejatest. Laines C2 oli 9 vastaja vanus puudu.

Andmed osalejate soolise koosseisu kohta igas küsitluslaines on tabelis 3.

Tabel 3: Osalejate sooline jaotus lainetes.

Sugu	Küsitluslaine									
	A1	A2	A3	A4	A5	A6	A7	C1	C2	
Naine	854 (55,2%)	725 (55,2%)	599 (53,9%)	500 (52,8%)	450 (55,9%)	361 (52,9%)	284 (49,1%)	185 (62,3%)	194 (64,9%)	
Mees	687 (44,4%)	584 (44,4%)	509 (45,8%)	444 (46,9%)	354 (44,0%)	322 (47,1%)	294 (50,8%)	112 (37,7%)	105 (35,1%)	
NA/muu	6 (0,4%)	5 (0,4%)	3 (0,3%)	3 (0,3%)	1 (0,1%)	-	1 (0,2%)	-	-	

Naiste arv on kuues täiskasvanute küsitluslainetes veidi suurem kui meeste arv. Naiste osakaal lainetes A1, A2, A3, A4, A5 ja A6 jääb vahemikku 53 kuni 55%. Ka alaealiste küsitluslainetes C1 ja C2 on naissoost osalejaid rohkem kui meessoost. Naissoost osalejad moodustavad mõlemas laines ligikaudu 63% kõigist osalejatest. Kõige väiksema osalusega täiskasvanute laines A7 on naisi veidi vähem kui mehi, naiste ja meeste osakaalud on siis peaaegu võrdsed (vastavalt 49% ja 51%).

Vastanute leibkonna suurused on tabelis 4 grupeeritud. Eraldi vaadatakse leibkondi, kuhu kuulub 1, 2, 3-5 või enam kui 5 liiget.

Tabel 4: Osalejate leibkondade suurused.

Leibkonna suurus	Küsitluslaine								
	A1	A2	A3	A4	A5	A6	A7	C1	C2
1	383 (24,8%)	320 (24,4%)	282 (25,4%)	249 (26,3%)	230 (28,6%)	197 (28,8%)	158 (27,3%)	-	-
2	664 (42,9%)	568 (43,2%)	489 (44,0%)	409 (43,2%)	358 (44,5%)	295 (43,2%)	244 (42,1%)	22 (7,4%)	19 (6,4%)
3-5	476 (30,8%)	405 (30,8%)	324 (29,2%)	277 (29,3%)	210 (26,1%)	184 (26,9%)	170 (29,4%)	256 (86,2%)	262 (87,6%)
6+	24 (1,6%)	21 (1,6%)	16 (1,4%)	12 (1,3%)	7 (0,9%)	7 (1,0%)	7 (1,2%)	19 (6,4%)	18 (6,0%)

Kõigis täiskasvanute küsitluslainetes on selliste osalejate, kes kuuluvad 2-liikmelisse leibkonda, osakaal suurim (enam kui 40%). Ühe- ja 3-5-liikmelistesse leibkondadesse kuulujaid on täiskasvanute küsitluslainetes peaaegu võrdselt. Lainetes A5 ja A6 on ühest inimesest koosneva leibkondi rohkem kui 3-5 inimesest koosneva leibkondi, ülejäänud täiskasvanute lainetes on üheliikmelisi leibkondi vähem kui 3-5-liikmelisi.

Valimisse ei saanud kuuluda üheliikmelistes leibkondades elavad alaealised. Enamik lastest (üle 85%) elab peredes, kus liikmeid on 3 kuni 5. Kui täiskasvanute lainetes jääb enam kui 5-liikmelistesse leibkondadesse kuulunud vastajate osakaal alati alla 2%, siis laste küsitluslainete puhul on see osakaal kolm korda kõrgem.

Andmed laste kooliskäimise kohta küsitlusele eelnenud päeval on toodud tabelis 5.

Tabel 5: Laste kooliskäimine küsitlusele eelnenud päeval vanusegrupiti.

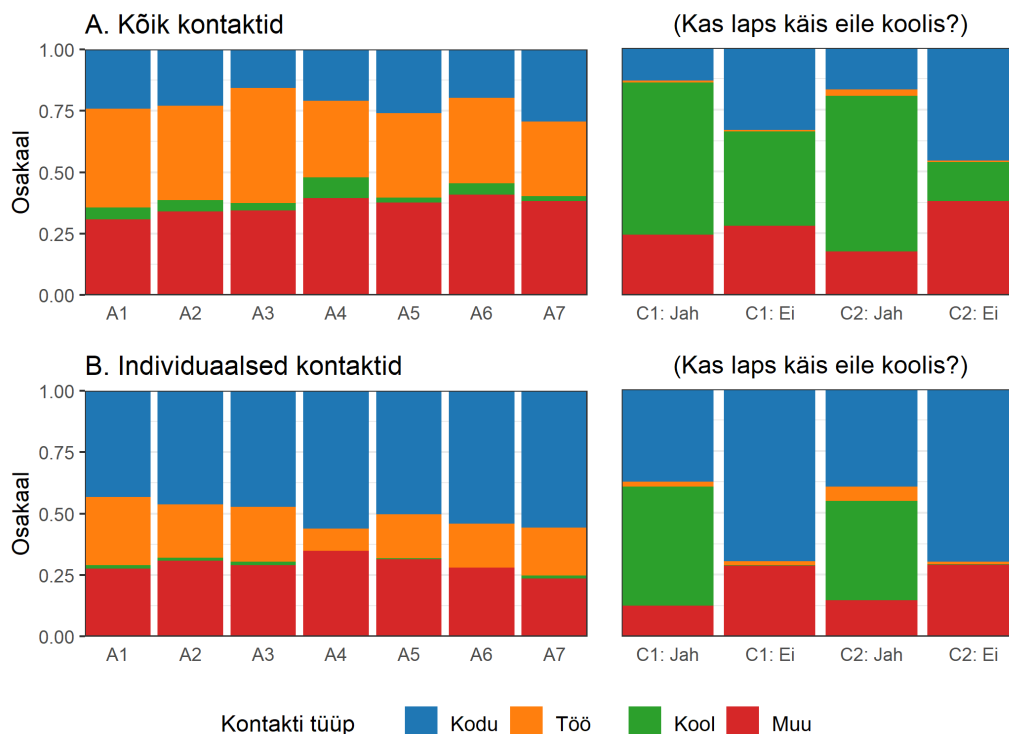
Kas laps käis koolis?	Küsitluslaine C1			Küsitluslaine C2		
	(0-17a)	(0-4a)	(5-17a)	(0-17a)	(0-4a)	(5-17a)
NA/Puudu	20 (6,7%)	16 (32,0%)	4 (1,6%)	21 (7,2%)	14 (25,0%)	7 (3,0%)
Ei	104 (35,0%)	22 (44,0%)	82 (33,2%)	241 (83,1%)	25 (44,6%)	216 (92,3%)
Ei, aga kool oli lapsele avatud	66 (22,2%)	14 (28,0%)	52 (21,1%)	25 (8,6%)	14 (25,0%)	11 (4,7%)
Ei, oli nädalalõpp/püha/vaba päev	29 (9,8%)	4 (8,0%)	25 (10,1%)	125 (43,1%)	6 (10,7%)	119 (50,9%)
Ei, kool oli suletud	7 (2,4%)	2 (4,0%)	5 (2,0%)	87 (30,0%)	4 (7,1%)	83 (35,5%)
Jah	175 (58,9%)	14 (28,0%)	161 (65,2%)	32 (11,0%)	18 (32,1%)	14 (6,0%)

Küsitluslaine C1 toimus juuni alguses ehk ajal, kui paljud haridusasutused olid veel avatud. 60% lastest käisidki küsitlusele vastamisele eelnenud päeval koolis. Suurele osale lastest (ligikaudu 80%) oli kool avatud.

Laine C2 toimus augustis ja kooliskäinute osakaal kõigist lastest oli vaid 11%. 5-kuni 17-aastaste seas oli kooliskäijaid 6%, noorematest (0-4a) käis kolmandik koolis (näiteks lasteaias või -sõimes). Kool oli avatud vähestele lastele, aga peamiseks põhjuseks ei ole koolide sulgemine viiruse tõkestamise eesmärgil, vaid asjaolu, et haridusasutused on üldjuhul augustis suvepuhkusel.

3.2 Kontaktide kirjeldus

Joonisel 2 on välja toodud kontaktide tüüpide jaotused küsitluslaineti. Küsitluslained C1 ja C2 on jagatud kaheks selle järgi, kas laps käis koolis või mitte. Eraldi vaadatakse kõiki osalejate poolt märgitud kontakte (kaasa arvatud need, mis on agregeeritult märgitud) ja kontakte, kus on arvestatud eraldi iga inimest.



Joonis 2: Kontaktide jaotus tüübi järgi, kui arvesse on võetud
 A. kõiki kontakte;
 B. kontakte, mille puhul on iga inimest eraldi arvestatud.

Küsitluslainetes A1 kuni A3 moodustavad tööl tehtud kontaktid kõigist kontaktidest ligikaudu 40%. Seda juhul, kui arvesse on võetud ka grupeeritult märgitud kontaktid. Töökontaktide osakaal langeb küsitluslaines A4, mille toimumisaeg (23. juuni kuni 3. juuli) viitab sellele, et osa töötajad on suvepuhkusele läinud. Lainetes A5 kuni A7 on töökontaktide osakaal püsivalt ligikaudu 30%. Kui arvestusest on grupeeritult märgitud kontaktid välja jäetud, siis töökontaktide osakaal on väiksem: kõigis küsitluslainetes 25% või veidi alla selle, välja arvatud laines A7, kus osakaal on ligikaudu 10%.

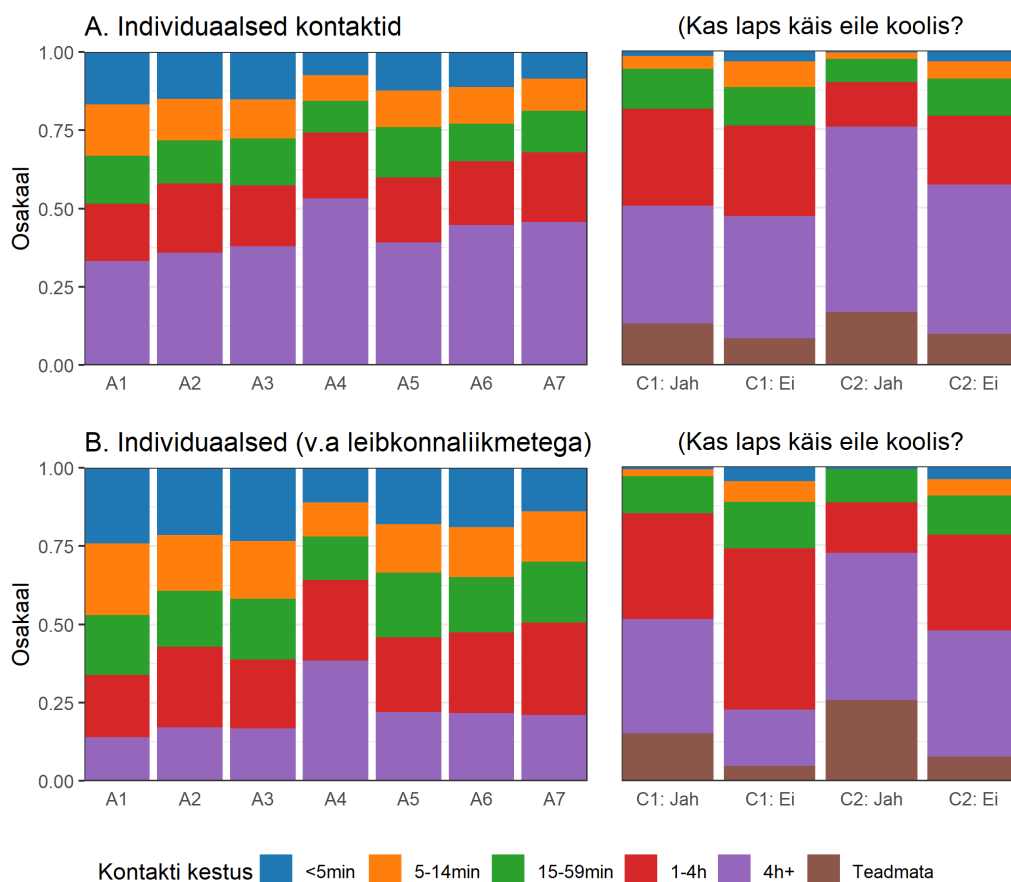
Koolikontakte märgitakse pigem grupeeritult, mitte ei arvestata eraldi igat inimest, kellega suheldi. Täiskasvanute lainetes on koolikontaktide osakaal väike ja püsib kogu uuringu käigus alla 8%. Kõigi kontaktide arvestuses on koolikontaktide osakaal

veidi suurem kui siis, kui grupeeritult märgitud kontaktid on välja jäetud.

Koduste kontaktide alla kuuluvad lisaks leibkonnaliikmete ja külalistega suhtlemisele ka kontaktid uksele, aias ja kodu juurdepääsude juures, näiteks trepid, liftid ja koridorid. Selliste kontaktide puhul võib olla lihtsam arvestada eraldi kõiki inimesi võrreldes teiste kontaktide tüüpidega. Kui arvesse on võetud agregeeritult märgitud kontaktid, siis lainetes A1, A5 ja A7 moodustavad kodused kontaktid 25% kõigist kontaktidest ning vahepealsetes lainetes on nende osakaal väiksem. Kui agregeeritult märgitud kontaktid on arvestusest välja jäetud, siis on nende osakaal suurem. Koduste kontaktide osakaal kasvab kuni küsitluslaineni A4, kui need moodustavad enam kui poole kõigist individuaalselt märgitud kontaktidest. Laines A5 nende osakaal veidi väheneb, aga edaspidi püsib ikkagi 50% juures.

Kontaktide jaotus tüübi järgi alaealiste küsitluslainetes C1 ja C2 on sarnane. Kui arvesse on võetud kõik märgitud kontaktid, siis laste puhul, kes käisid koolis, on enam kui 60% tehtud kontaktidest kooliga seotud ja ligikaudu 15% kontaktidest on tehtud kodus. Kui välja arvata grupeeritult märgitud kontaktid, siis koolis käinute koolikontaktid moodustavad siiski kõige suurema osa tehtud kontaktidest (veidi alla 50%), aga koduste kontaktide osakaal on suurem (enam kui 30%). Nii lainetes C1 ja C2 on laste puhul, kes ei käinud koolis, märgitud kooliga seonduvaid kontakte ainult grupeeritult. Laines C1 on selliseid kontakte kaks korda rohkem kui laines C2. Töökontakte on lastel väga vähe.

Kontaktide, kus iga inimest on eraldi arvestatud, jaotus kestuse järgi küsimuslaineti on kujutatud joonisel 3. Küsitluslained C1 ja C2 on jagatud kaheks selle järgi, kas laps käis koolis või mitte. Eraldi vaadatakse kõiki individuaalselt märgitud kontakte ja individuaalseid kontakte, mille seast on kontaktid leibkonnaliikmetega välja jäetud.



Joonis 3: Kontaktide jaotus kestuse järgi, kui arvesse on võetud
 A. individuaalseid kontakte;
 B. individuaalseid kontakte, mille seast on kontaktid leibkonnaliikmetega välja arvatud.

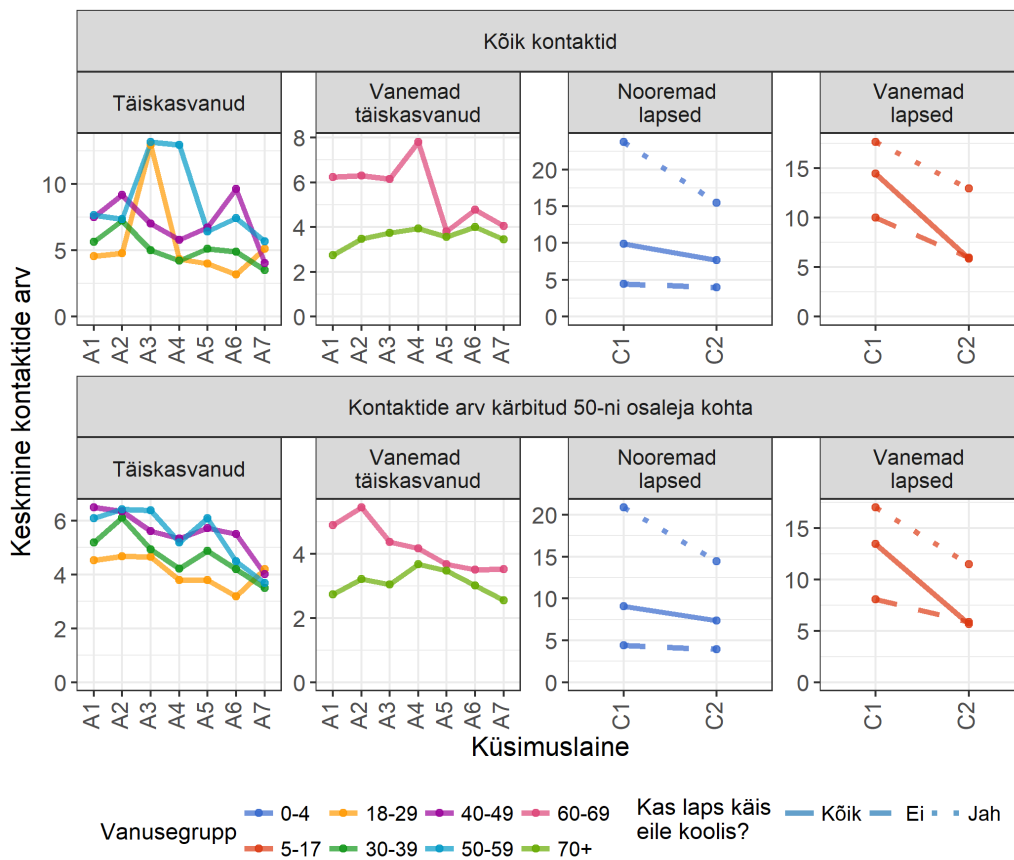
Kõige suurema osa kõigist individuaalsetest kontaktidest moodustavad täiskasvanute küsitluslainetes üle 4 tunni kestvad kontaktid. Suurim on nende osakaal küsitluslainetes A4 ja A7. 1 kuni 4 tundi kestvad kontaktid moodustavad kõigis täiskasvanute küsitluslainetes ligikaudu 20%, 15-59 minutit kestvate kontaktide osakaal on väiksem. Vähem kui 15 minutit kestvad kontaktid moodustavad küsitluslaines A1 ligikaudu 30%, aga järgnevatel lainetes nende osakaal langeb. Vähim on lühikeste kontaktide osakaal lainetes A4 ja A5, ligikaudu 20%.

Leibkonnaliikmetega suheldakse päevas kaua ehk kui kontaktid nendega välja ar-

vestada, siis üle 4 tunni kestvate kontaktide osakaal on kõigi küsitluslainete puhul madalam. Jättes välja kontaktid leibkonnaliikmetega, on kontaktide jaotus kestuse järgi täiskasvanute lainetes ühtlasem. Laine A4 paistab silma veidi teistsugusema jaotuse poolest võrreldes temale eelneva ja järgneva küsitluslainega. Kauem kestvate kontaktide osakaal on seal suurem ja alla tunni kestvate kontaktide osakaal on teiste lainetega võrreldes kõige väiksem. Kuna selles küsitluslaines muutus ka töökontaktide osakaal väiksemaks, võib oletada, et suvepuhkusel on jäänud vähemaks lühemad vestlused töökaaslastega ja rohkem suheldakse tuttavatega, kes on lähedasemad ja kellega kontaktid kestavad kauem.

Alaealiste küsitluslainetes C1 ja C2 on kõigi individuaalsete kontaktide arvestuses koolis käinute ja mittekäinute puhul kontaktide jaotus kestuse järgi sarnane. Üle tunni kestvad kontaktid moodustavad ligikaudu 80% kõigist kontaktidest ja vähem aega kestvate kontaktide osakaal on väike. Kui kontaktid leibkonnaliikmetega arvestusest välja jätta, siis üle tunni kestvate kontaktide osakaal on ikkagi ligikaudu 80%, aga küsitluslaine C1 puhul väheneb üle 4 tunni kestvate kontaktide osakaal märgatavalt (nii koolis käinutel kui ka mittekäinutel). Küsitluslaines C2 muutus nii suur ei ole. See tähendab, et suve lõpus on lastel kaua kestvaid kontakte tehtud rohkem inimestega, kes nende leibkonda ei kuulu. Lastel on alla 5 minuti kestvaid (võibolla pinnapealsemaid) kontakte väga vähe, nende osakaal jääb alati alla 5%.

Joonisel 3 on kujutatud keskmise kontaktide arvu muutus kogu uuringu käigus vanusegrupiti. Kuna andmestikus on igas küsitluslaines inimesi, kes on kontakte märkinud eriliselt palju (küsitluslaines A1 ja A2 on maksimaalne kontaktide arv ligikaudu 400, küsitluslaines A3 aga 1300), siis erindite mõju vähendamiseks on kontaktide arvu maksimum tõkestatud arvuga 50. See tähendab, et osalejate puhul, kes märkisid kontakte enam kui 50, läheb kontaktide arvuna arvesse täpselt 50. Eraldi vaadatakse keskmise kontaktide arvu muutumist kõigi märgitud kontaktide ja ka kärbitud kontaktide korral.



Joonis 4: Keskmine kontaktide arv küsimuslainetes vanusegrupiti. (Esimesel juhul arvesse võetud kõik märgitud kontaktid, teisel juhul on kontaktide arv kärbitud 50-ni osaleja kohta.)

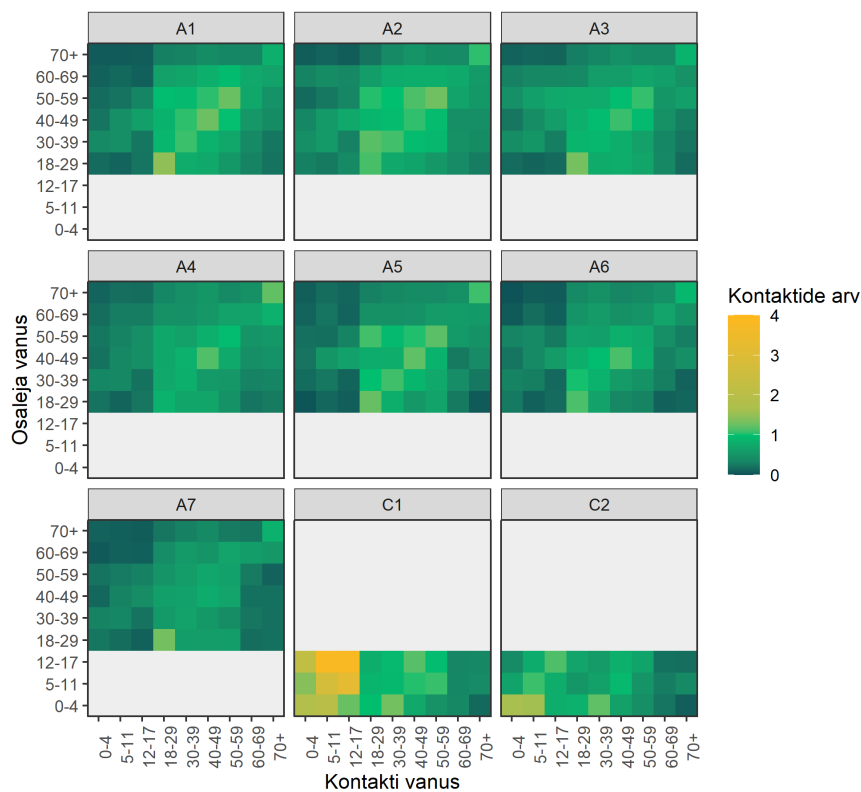
Kui kontaktide arv on kärbitud arvuni 50 iga osaleja kohta, siis keskmine kontaktide arv järjest langeb uuringu käigus peaaegu kõigis vanusegruppides. Teistest erineb 70-aastaste ja vanemate vanusegrupp, kus küsitluslaineni A4 keskmine kontaktide arv kasvab, aga seejärel hakkab langema ning uuringu lõpus ehk küsitluslaines A7 on peaaegu võrdne sellega, mis uuringu alguses.

Küsitluslainetes C1 ja C2 on nii nooremate kui ka vanemate laste puhul kooliskäivate keskmine kontaktide arv kõrgem kui neil, kes koolis ei käinud. Suve alguses (laine C1) on keskmine kontaktide arv lastel kõrgem kui suve lõpus (laine C2), olenemata sellest, kas käidi küsitlusele eelneval päeval koolis või mitte.

Kõigi kontaktide arvestuses on näha, et üksikud väga suured ööpäevased kontaktide arvud mõjutavad keskmist arvu tugevalt küsitluslainetes A3, A4 ja ka A5. Mõju on näha 18- kuni 29-aastaste, 50- kuni 59-aastaste ja ka 40- kuni 49-aastaste seas. Neis vanusegruppides keskmine kontaktide arv tõuseb ja langeb uuringu käigus järsemalt kui teistes vanusegruppides.

3.3 Kontaktimaatriksid

Kontaktimaatriksite loomisel on osalejate ööpäevane kontaktide arv kärbitud arvuni 50. Esiteks on bootstrap-meetodil loodud igale küsitluslainele vastav ebasümmeetriline (lünklik) kontaktimaatriks (vt Joonis 5).

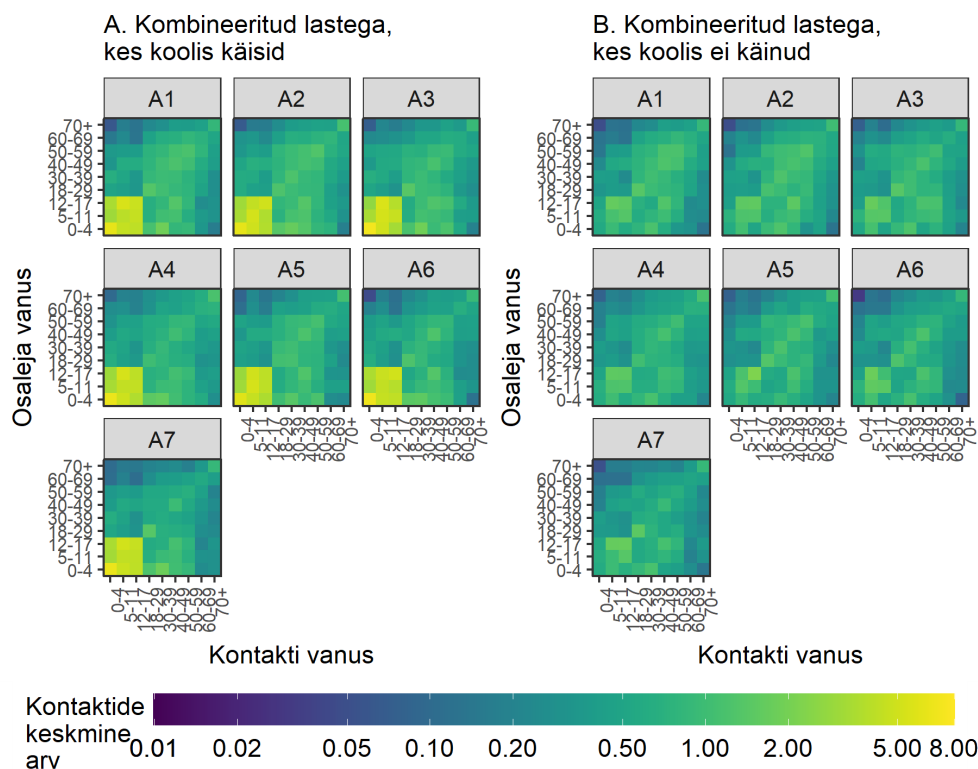


Joonis 5: Kontaktimaatriksid iga küsitluslaine kohta eraldi (kontaktide arv kärbitud 50-ni osaleja kohta).

Kõige enam suheldakse oma vanusegrupist või sellele lähedal paiknevast vanuse-

grupist inimestega. Sellele viitab maatriksi eredam kõrvaldiagonaal (erksam värv tähistab kõrgemat kontaktide arvu antud vanusegruppide vahel). Silma paistab juunis toimunud alaealiste küsitluslaine C1, kui 5-17-aastased puutusid tihedalt kokku oma vanusegrupi lastega. Põhjus seisneb ilmselt selles, et suur osa neist (65%, vt Tabel 5) käisid küsitlusele eelneval päeval mingisuguses haridusasutuses ehk said seal suhelda palju eakaaslastega. Alaealiste osalejate puhul (nii küsitluslained C1 ja C2) on heledalt tähistatud ka kontaktisikute vanusegrupid 30-39 ja 40-49, kuhu arvatavasti kuuluvad laste vanemad või nende eest hoolt kandvad inimesed.

Kombineerides täiskasvanute küsitluslaineid alaealiste küsitluslainetega, saadi teraviklikud kontaktimaatriksid. Joonisel 6 on täiskasvanute lained kombineeritud eraldi koolis käinute ja nendega, kes koolis ei käinud (mõlemast lainest C1 ja C2).

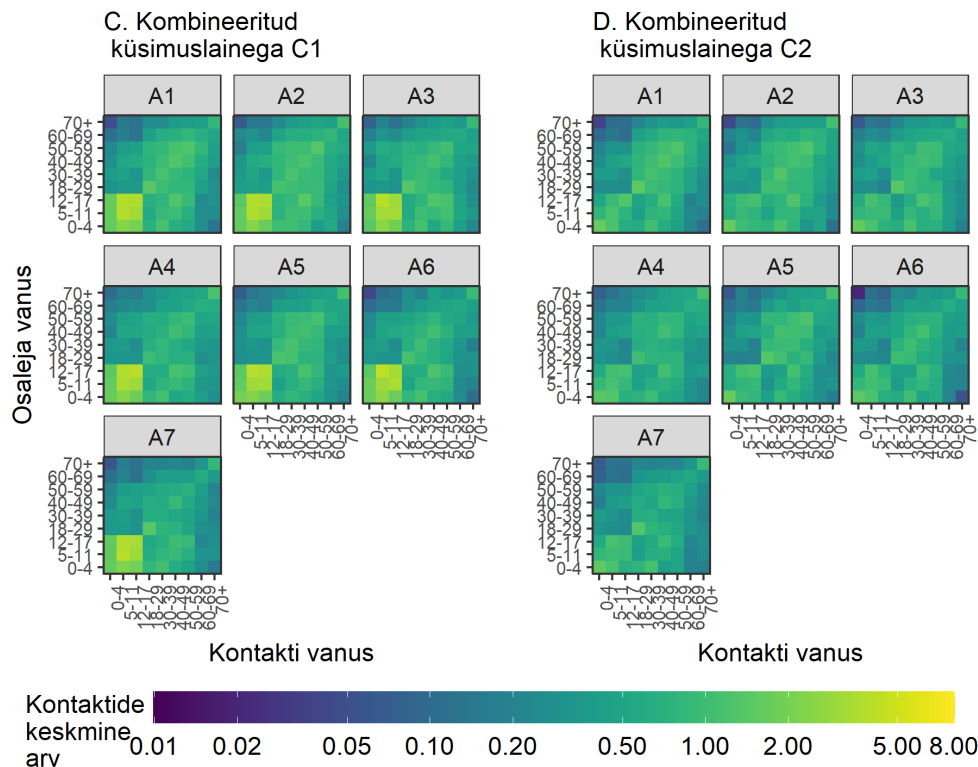


Joonis 6: Kontaktimaatriksid, kui täiskasvanute kontaktid on kombineeritud A. koolis käinud laste kontaktidega; B. koolis mitte käinud laste kontaktidega.

Kombineerides täiskasvanute kontakte koolis käinute kontaktidega, on kõigi küsitluslainete kontaktimaatriksite all vasakus nurgas kollane ruut, mis vastab keskmiselt suuremale kontaktide arvule. Lastel on koolis eakaaslastega palju kontakte olnud, keskmiselt 5 kuni 8 kontakti ööpäevas. Teine veidi heledam ruut asub 18- kuni 59-aastaste vahel. See tähendab, et tööealised suhtlevad veidi rohkem teiste tööealistega. Erinevate lainete maatriksite vahel suuri erinevusi pole. Lainete A1 kuni A4 maatriksid on väga sarnased, edasistes lainetes hakkavad maatriksid (väljastpoolt koolilaste kollast ruutu) veidi tuhmimaks muutuma ehk kontaktide arv väheneb.

Lastel, kes koolis ei käinud, on olnud vähem kontakte eakaaslastega, aga täiskasvanutega on nad suhelnud sarnaselt nende lastega, kes koolis käisid. Seega kui täiskasvanute kontaktid on kombineeritud koolis mitte käinud laste kontaktidega, siis tekkinud maatriksid on sarnased koolilaste kontaktidega kombineerides saadud maatriksitega. Erinev on ainult ruut, mis tähistab kontakte lastel oma eakaaslastega.

Joonisel 7 on toodud kontaktimaatriksid, kus ühel juhul on täiskasvanute lainetes märgitud kontaktid kombineeritud laine C1 kontaktidega, ja teisel juhul, laine C2 kontaktidega. Lainega C1 kombineeritud maatriksites on vasakul all nurgas heledam ruut, sest suur osa küsitluslainest C1 osavõtjatest käis koolis. Maatriksites, mis on kombineeritud laine C2 kontaktidega, on see ruut küll näha, aga selle värvus on tumedam. Ülejäänud osa maatriksitest sarnaneb joonisel 6 kujutatud maatriksitega.



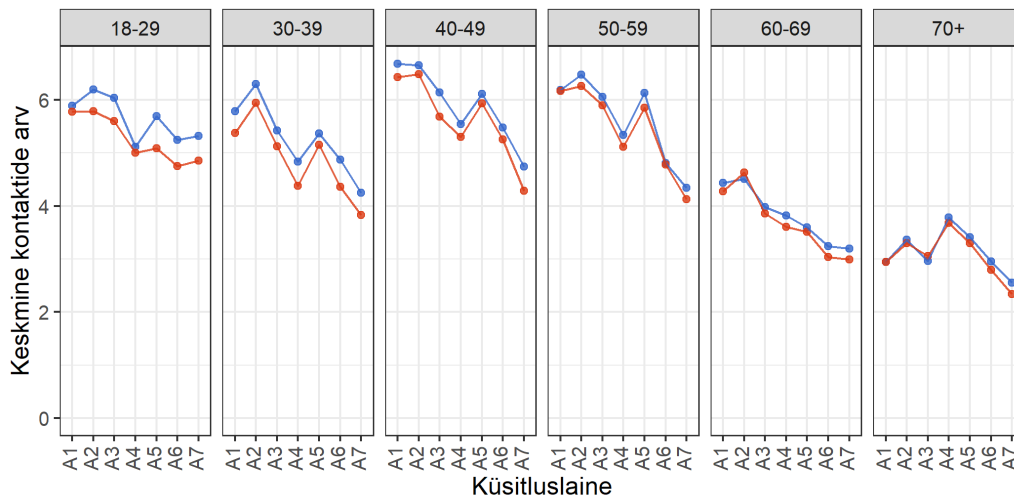
Joonis 7: Kontaktimaatriksid, kui täiskasvanute kontaktid on kombineeritud C. küsitluslaine C1 kontaktidega; D. küsitluslaine C2 kontaktidega.

Joonisel 8 on kujutatud keskmist kontaktide arvu vanusegrupiti, mis on leitud samal põhimõttel nagu sümmeetrilised kontaktimaatriksid. Vaadatakse eraldi juhte, kus täiskasvanute kontaktid on esmalt kombineeritud koolis käinud laste kontaktidega ja seejärel koolis mitte käinud laste kontaktidega.

Joondiagrammid vanusegruppides 18-29, 30-39, 40-49 ja 50-59 näevad sarnased välja. Uuringu alguses, küsitluslainetes A1 ja A2 on keskmine kontaktide arv suurim ning uuringu käigus see langeb. Laines A5 korraks tõuseb keskmine kontaktide arv kõigis vanusegruppides, aga langeb laines A6 samale tasemele, mis oli laines A4, või veel madalamale. Koolis käinud laste kontaktidega kombineerides leitud keskmine kontaktide arv on veidi kõrgem kui koolis mitte käinud laste kontaktidega kombineerides, aga vahe pole suur, ei ületa kunagi 0,5 kontakti. 50-aastaste ja

vanemate puhul ei ole erinevust enam üldse näha. See kinnitab tähelepanekut, et need, kes koolis käisid, ja need, kes ei käinud, on täiskasvanutega suhelnud peaaegu sama palju.

60-aastaste ja vanemate puhul on näha, et kontaktide keskmine arv on madalam kui noorematel kogu uuringu käigus. Vanusegrupis 60-69 langeb keskmine kontaktide arv uuringu jooksul neljalt kolmeni. 70-aastaste ja vanemate seas on keskmine kontaktide arv laines A7 veidi madalam kui laines A1, aga vahepealsetes lainetes on see kõrgem võrreldes uuringu alguse ja lõpuga. Maksimaalne keskmine kontaktide arv saavutatakse küsitluslaines A4.



Kombineeritud lastega, kes — käisid koolis — ei käinud koolis

Joonis 8: Keskmine kontaktide arv täiskasvanute vanusegruppides, kui nende kontaktid on kombineeritud

- A. koolis käinud laste kontaktidega;
- B. koolis mitte käinud laste kontaktidega.

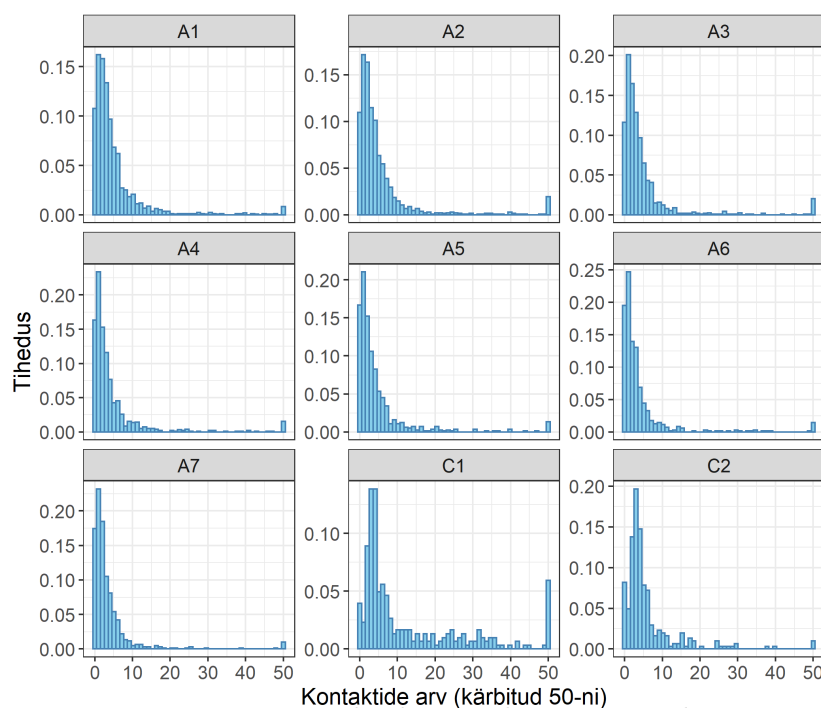
3.4 Regressioonimudel päevas tehtud kontaktide arvule

Ühes ööpäevas tehtud kontaktide arvule hinnati negatiivse binoomregressiooni mudel. Kuna ühes küsitluslaines osaleja sai osaleda ka teistes, siis on töös hinnatud mudeleid iga küsitluslaine jaoks eraldi. Andmeid saaks kombineerida segamudeleid

kasutades, kuid see oleks tehniliselt keerulisem, ning on plaanis edaspidi.

Joonisel 9 on kujutatud osalejate ööpäevas tehtud kontaktide arvude histogrammid iga küsitluslaine kohta. Kontaktide arvud on eelnevalt kärbitud arvuni 50. See tähendab, et kõigi puhul, kes olid märkinud kontakte enam kui 50, märgitakse kontaktide arvuks täpselt 50.

Kõigi täiskasvanute lainete histogrammid on sarnase kujuga. Kontaktide arvu jaotused on paremale kaldu, jaotustel on pikk saba paremal ehk suurte väärtuste pool. Kontaktide arvu 50 juures on näha vasakul paiknevatest tulpadest veidi kõrgem tulp, mille põhjuseks on kontaktide arvu kärpimine. Kõige rohkem on igas küsitluslaines inimesi, kes on ööpäevas olnud vahetus kontaktis ühe isikuga.



Alaealiste küsitluslainete histogrammid erinevad täiskasvanute lainete omadest. Vähenenud osalejate arv võib olla põhjuseks, miks histogrammid ei näe nii korrapäraseid välju. Laines C1 esinevad kõige sagedamini kontaktide arvu väärtused kolm ja neli (mõlemat 42 korda), laines C2 on kõige sagedasem kontaktide arvu väärtus

kolm (60 korda). Laines C1 on 18 osaleja puhul märgitud enam kui 49 kontakti ehk tulp kontaktide arvu 50 kohal on palju kõrgem, kui vasakule jäävad tulbad. Kuna laste kontaktide arvude histogrammid on täiskasvanute lainete histogrammi-dest erinevad ja laste lainetes on vähem osalejaid, siis on töös mudelid tehtud vaid täiskasvanute küsitluslainetes osalenute kontaktide arvule.

Tunnusteks, mis võiksid ennustada ööpäevas tehtud kontaktide arvu, valiti 2008. aastal avaldatud kontaktide uuringu POLYMOD andmete analüüsi eeskujul (2008) osaleja vanus, sugu ja leibkonna suurus. Leibkonna suurus pandi mudelisse 4-tasemelise faktortunnusena, mille tasemeteks määrati 1, 2, 3 ja enam kui 3 liikmega leibkonnad.

Küsitluslaine A1 puhul hinnatud mudel on järgmine:

$$\begin{aligned} \ln \mu = & 1,40 - 0,001 \cdot \text{vanus} + 0,13 \cdot I_{(\text{sugu}=\text{mees})} + \\ & + 0,09 \cdot I_{(\text{leibkonna suurus}=2)} + 0,45 \cdot I_{(\text{leibkonna suurus}=3)} + \\ & + 0,60 \cdot I_{(\text{leibkonna suurus}>3)}, \end{aligned}$$

kus μ on keskmine kontaktide arv ja I tähistab indikaatortunnust. See tähendab, et $I_{(\text{sugu}=\text{mees})}$ väärtus on 1, kui mudeldatakse meessoost isiku kontaktide arvu, ja 0 muidu.

Osutus, et üheski küsitluslaine mudelis ei olnud vanus olulisuse nivool 0,05 oluline tunnus. See tähendab, et täiskasvanute lainetes, kus osalejad olid vähemalt 18-aastased, ei mõjutanud vanus ööpäevas tehtud kontaktide arvu. Selgus ka, et sugu tuli statistiliselt oluline ainult küsitluslaines A1, ning otsustati, et sugu jäetakse välja kõigi küsitluslainete mudelitest. Laine A1 lõplik mudel on kujul:

$$\begin{aligned} \ln \mu = & 1,36 + 0,11 \cdot I_{(\text{leibkonna suurus}=2)} + \\ & + 0,48 \cdot I_{(\text{leibkonna suurus}=3)} + 0,65 \cdot I_{(\text{leibkonna suurus}>3)}. \end{aligned}$$

Eeldusel, et kontaktide arvu suhtes on tegemist juhuvalimiga, saab mudeli abil

hinnata erinevate suurustega leibkondadesse kuulunud inimeste keskmist ööpäevas tehtud kontaktide arvu. Hinnang keskmisele kontaktide arvule avaldub järgmiselt:

$$\mu = \exp(1,36 + 0,11 \cdot I_{(\text{leibkonna suurus}=2)} + 0,48 \cdot I_{(\text{leibkonna suurus}=3)} + 0,65 \cdot I_{(\text{leibkonna suurus}>3)}).$$

Seega 3-liikmelises leibkonnas oli inimestel küsitluslaine A1 toimumise perioodil keskmiselt $\exp(0,48) \approx 1,62$ korda rohkem kontakte kui üheliikmelises. Tabelis 6 on toodud leibkonna suuruse multiplikatiivne mõju kontaktide arvule. Väärtus lahtris näitab, mitu korda erines kontaktide arv antud suurusega leibkonnas kontaktide arvust üheliikmelises leibkonnas. Iga väärtuse järel on 95%-usaldusintervall. Kui intervalli kuulub arv 1, siis selles laines ei erinenud kontaktide arv vaadeldava suurusega leibkonnas oluliselt üheliikmelises leibkonnas esinenud kontaktide arvust.

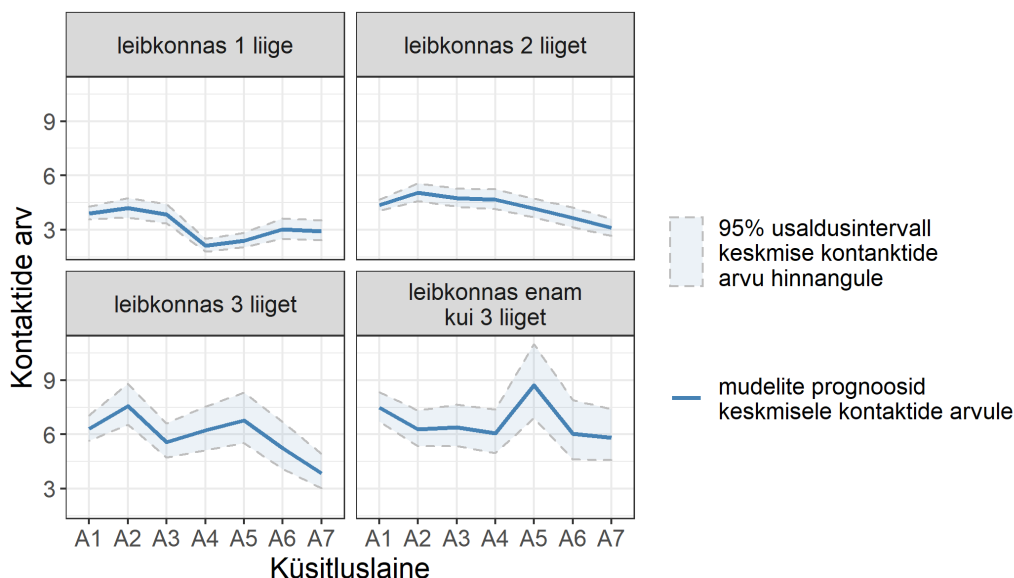
Tabel 6: Leibkonna suuruse multiplikatiivne mõju kontaktide arvule koos 95%-usaldusintervalliga (võrdlusgrupp: leibkond suurusega 1).

Küsitluslaine	Leibkonna suurus		
	2	3	> 3
A1	1,11 (0,97; 1,28)	1,62 (1,36; 1,92)	1,91 (1,62; 2,27)
A2	1,21 (1,03; 1,42)	1,81 (1,49; 2,21)	1,50 (1,23; 1,84)
A3	1,24 (1,04; 1,47)	1,45 (1,17; 1,81)	1,67 (1,33; 2,10)
A4	2,20 (1,80; 2,70)	2,95 (2,30; 3,80)	2,86 (2,22; 3,71)
A5	1,74 (1,42; 2,14)	2,83 (2,18; 3,68)	3,64 (2,75; 4,86)
A6	1,21 (0,95; 1,53)	1,73 (1,28; 2,37)	2,00 (1,45; 2,79)
A7	1,06 (0,83; 1,35)	1,32 (0,97; 1,81)	2,00 (1,48; 2,72)

Kontaktide arvu poolest oli kõige suurem erinevus ühe- ja 2-liikmeliste leibkondade vahel lainetes A4 ja A5. Ülejäänud lainetes oli erinevus pigem väike või isegi statistiliselt ebaoluline. Laines A7 erinesid oluliselt vaid väiksematesse leibkondadesse (3 liiget või alla selle) ja suurematesse leibkondadesse (enam kui 3 liiget) kuulunud isikute keskmine kontaktide arv.

Joonisel 10 on kujutatud erinevate leibkonna suuruste korral hinnang keskmisele

kontaktide arvule kõigis küsitluslainetes koos 95% usalduspiiridega.



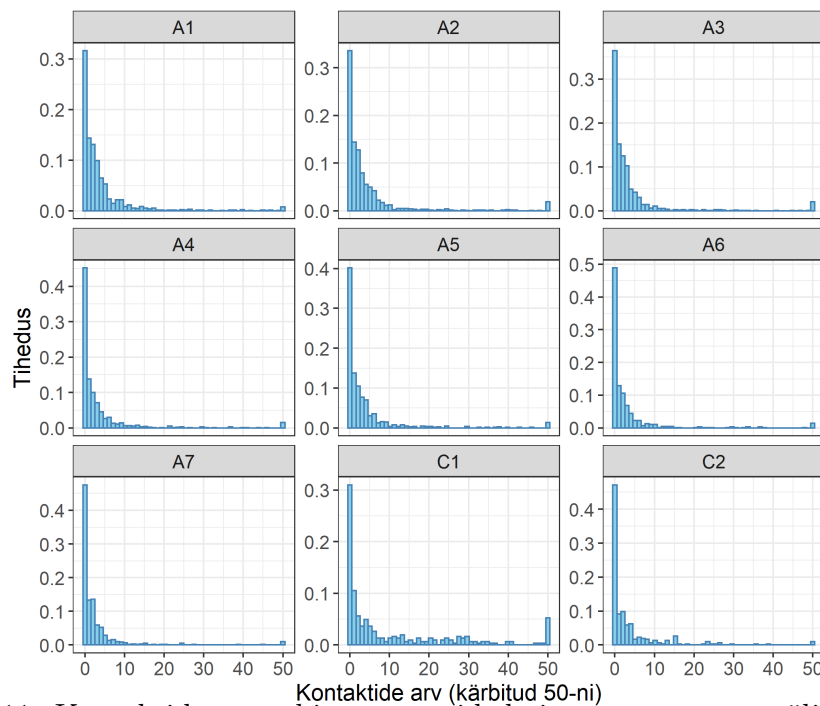
Joonis 10: Kõigi küsitluslainete mudelite prognoosid keskmisele kontaktide arvule koos 95% usaldusintervallidega.

Ühe- ja 2-liikmeliste leibkondade puhul on keskmise kontaktide arvu hinnangutel kitsad usalduspiirid. Keskmise kontaktide arv on neis leibkondades uuringu algusest ehk lainest A1 kuni laineni A3 sarnasel tasemel. Üheliikmelised leibkonnad on lainetes A4 ja A5 kontaktide arvu märgatavalt langetanud, kuid suuremates leibkondades seda toimunud pole. 2-liikmelistes leibkondades langeb keskmine lainest A2 alates sujuvalt ja saavutab uuringu lõpuks sarnase taseme, mis üheliikmelistes leibkondadeski.

Suuremates leibkondades on usaldusvahemikud keskmisele kontaktide arvule laiemad. 3-liikmelistes leibkondades keskmise kontaktide arvu usaldusvahemikud uuringu alguses ja uuringu lõpus ei kattu ehk keskmine kontaktide arv uuringu käigus on langenud. Suuremates leibkondades seda väita ei saa, sest laine A1 ja laine A7 mudelite usaldusintervallid kattuvad. Leibkondades, kus on enam kui 3 liiget püsib keskmine kontaktide arv uuringu käigus samal tasemel, välja arvatud küsitluslaines A5, kus keskmine tõuseb 9 kontaktini ööpäevas. Tõenäosusega 95% langeb suurde

leibkonda kuuluvate inimeste keskmine kontaktide arv küsitluslaine A5 toimumise perioodil vahemikku 7 kuni 11 kontakti ööpäevas.

Ööpäevastele kontaktide arvule loodud mudelid aitaksid uurida, kuidas piisnakkusena leviva viiruse tõkestamiseks rakendatud meetmed kontaktide arvu mõjutanud on. Sel juhul tuleks arvestusest välja jätta leibkonnaliikmetega aset leidnud vahetud kontaktid, sest nende arvu piirangud ei vähenda. Kontaktide arvu histogrammid ilma leibkonnasisestest kontaktideta on toodud joonisel 11. Nüüd ei erine laste küsitluslainete histogrammid täiskasvanute lainete omadest enam nii märgatavalt. Kõigis lainetes on kõige enam neid, kellel polnud ühtegi vahetut kontakti oma leibkonda mitte kuuluva inimesega.



Joonis 11: Kontaktide arvu histogrammid, kui arvestusest on välja jäetud leibkonnasisesed kontaktid.

Mudeleid leibkonnavälisetele kontaktide arvule ehitati eelnevalt kirjeldatud viisil. Osutus, et sellist argumenttunnust, mis oleks suuremas osas lainetes oluline kontaktide arvu prognoosimisel, polnud. Keskmised kontaktide arvud erinesid oluliselt

erinevate suurusetega leibkondades vaid küsitluslainetes A1 ja A4. Vanusel ja sool polnud olulist mõju kontaktide arvule üheski täiskasvanute küsitluslaines. Alaealiste laines C1 oli sool mõju kontaktide arvule, poisid suhtlesid rohkem kui tüdrukud. Laines C2 seevastu ei erinenud poiste ja tüdrukute keskmine kontaktide arv oluliselt.

Kokkuvõte

Töö eesmärk oli uurida seaduspärasusi eestlaste suhtlemises ja kirjeldada nende muutumist uuringu jooksul. Üldiselt kõigis vanusegruppides keskmine kontaktide arv vähenes. Silma paistsid küsitluslained A4 ja A5 ehk periood 23. juunist 22. juulini. Kontaktide omaduste muutumist ja nende keskmise arvu taseme kõikumist sellel perioodil mõjutab arvatavasti tööealiste suvepuhkusel olemine.

Kontaktimaatriksitelt võis välja lugeda, et inimesed suhtlevad rohkem eakaaslastega. Selgelt eristus koolis käinud ja mitte käinud laste suhtlemine. Kuigi 2021. aasta suvel ei olnud koolid üldiselt koroonaviiruse tõkestamise eesmärgil suletud, oli töös võimalik võrrelda perioodi, kui suur osa koolidest on avatud, perioodiga, kui haridusasutused on suletud. Nii koolis käinud ja mitte käinud alaealised suhtlesid täiskasvanutega sama palju, aga koolis käinutel oli päevas palju rohkem kontakte eakaaslastega.

Kõigi kontaktide arvestuses ehitati mudelid kontaktide arvule vaid täiskasvanute lainetes. Selgus, et 18-aastaste ja vanemate puhul ei mõjutanud sugu ja vanus ööpäevas tehtavate kontaktide arvu. Oluliselt mõjutas kontaktide arvu vaid vastaja leibkonna suurus. Mida suuremasse leibkonda inimene kuulub, seda rohkem on tal ööpäevas vahetuid kontakte. 1-, 2- ja 3-liikmeliste leibkondadesse kuulunute keskmine kontaktide arv langes uuringu jooksul. Suurematesse leibkondadesse kuulunute keskmine kontaktide arv võis usalduspiire arvestades püsida samal tasemel kogu uuringu käigus. Kuna viiruse tõkestamiseks rakendatavad meetmed ei piiraks leibkonnasiseseid kontakte, oleks kasulikum uurida mudeleid ööpäevas tehtud kontaktide arvule, kuhu hulka pole loetud leibkonnaliikmetega toimunud kontakte. Leiti aga, et abitunnuseid, mis aitaksid hinnata suuremas osas küsitluslainetes leibkonnaväliste kontaktide arvu, ei olnud.

Kasutatud allikad

- Eesti Statistikaamet (kuupäev puudub). *Rahvastikupüramiid*. URL: <https://www.stat.ee/rahvastikupyramiid/> (vaadatud 17.03.2022).
- Funk, Sebastian (2020). *socialmixr: Social Mixing Matrices for Infectious Disease Modelling*. R package version 0.1.8. URL: <https://CRAN.R-project.org/package=socialmixr>.
- Gimma, Amy, Kerry Wong, Kevin Van Zandvoort, Christopher Jarvis ja John Edmunds (2022). *Comix Epipose*. https://github.com/amygimma/comix_epipose.
- Hilbe, Joseph M. (2011). *Negative binomial regression / Joseph M. Hilbe*. eng. Second edition. Cambridge: Cambridge University Press. ISBN: 1-107-21504-8.
- Mossong, Joël, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga ja others (2008). "Social contacts and mixing patterns relevant to the spread of infectious diseases". *PLoS medicine* 5.3. Publisher: Public Library of Science San Francisco, USA, e74.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Hanna Sõnajalg,

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

Hanna Sõnajalg

10.05.2022