

Tartu Ülikool  
Loodus- ja täppisteaduste valdkond  
Matemaatika ja statistika instituut

Tuule Tamme

**Radioloogiliste korduvuuringute tuvastamine  
globaalsel skaalal Eesti Haigekassa näitel**

Matemaatilise statistika eriala  
Bakalaureusetöö (9 EAP)

Juhendaja Sven Laur, DSc. (Tech)

Tartu 2019

# **Radioloogiliste korduvuuringute tuvastamine globaalsel skaalal Eesti**

## **Haigekassa näitel**

Bakalaureusetöö

Tuule Tamme

**Lühikokkuvõte.** Käesoleva bakalaureusetöö eesmärk on kirjeldada radioloogiliste uuringute anomaaliaid. Töö esimeses peatükis kirjeldatakse radioloogiliste uuringute ajalisi anomaaliaid võrreldes üldise teenusekasutusega Eestis. Analüüs viiakse läbi Eesti Haigekassa raviarvete baasil. Teise peatüki eesmärk on tuvastada korduvate radioloogiliste uuringute tegemist.

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

**Märksõnad:** Klasteranalüüs, binoomtest

# **Identification of repeated radiological examinations on a global scale based on Estonian Health Insurance Fund**

Bachelor's thesis

Tuule Tamme

**Abstract.** The purpose of this bachelor thesis is to describe the anomalies of radiological studies. The first part of the thesis describes the temporal anomalies of radiological studies compared to the general use of services in Estonia. The analysis is carried out on the basis of the EHIF's medical invoices. The goal in the second part of the thesis is to identify repeated radiological studies.

**CERCS research specialisation:** P160 - Statistics, operations research, programming, actuarial mathematics

**Keywords:** Cluster analysis, binomial test

# Sisukord

<b>Sissejuhatus</b>	<b>2</b>
<b>1 Metoodika</b>	<b>4</b>
1.1 Binoomtest . . . . .	4
1.2 Keskmise absoluutne viga . . . . .	4
1.3 Hierarhiline klasterdamine . . . . .	4
<b>2 Globaalse teenusekasutuse anomaaliate otsing</b>	<b>6</b>
2.1 Andmestiku ülevaade . . . . .	6
2.2 Üldmudeli fikseerimine . . . . .	6
2.3 Radioloogilise teenusekasutuse võrdlus globaalse teenusekasutusega .	12
2.4 Radioloogiliste uuringute teenusekasutus aastate lõikes . . . . .	14
2.5 Radioloogiliste uuringute klasterdamine . . . . .	21
2.6 Järeldus . . . . .	34
<b>3 Korduvate uuringute tuvastamine</b>	<b>36</b>
3.1 Andmestiku ülevaade . . . . .	36
3.2 Korduvate uuringute tuvastamine . . . . .	36
<b>Kokkuvõte</b>	<b>45</b>
<b>Kasutatud kirjandus</b>	<b>47</b>

# Sissejuhatus

Eestis korraldab riiklikku ravikindlustust Eesti Haigekassa. Eesti Haigekassale esitatakse raviarved tehtud uuringute kohta ning antud andmete põhjal on võimalik uurida teenusekasutust. Töös uuritakse radioloogiliste uuringute teenusekasutust. Radioloogilised uuringud on röntgen, ultraheli, kompuutertomograafia ja magnetresonantstomograafia. Levinud probleemiks on patsientidele korduvate radioloogiliste uuringute tegemine. Varasemalt on antud probleemi kohta maailmas tehtud mitmeid uuringuid (James Nol 2005). Teatavasti annavad radioloogilised uuringud kiirgusdoose, mis korduvate uuringute korral võivad patsiendile kahjulikud olla. Seetõttu soovitakse teada, kas sama probleem võib esineda ka Eestis.

Töö esimeses osas uuritakse teenusekasutuse ajalisi anomaaliaid, mis jagunevad kaheks. Teenusekasutuses võivad olla nii ajaliselt püsivad, kui ka ajutised anomaaliad. Vahetevahel võivad radioloogilise uuringu valikut erinevate teenuste vahel mõjutada ravivälised faktorid. Näiteks otsustatakse röntgenuuringu asemel teha kompuutertomograafia uuring, kuigi mõistlikum oleks kasutada röntgenuuringut. Kui selliseid otsuseid on piisavalt palju, siis on need üldises teenusekasutuses tuvastatavad püsivate anomaaliatena. Töö esimeses osas proovitakse leida meetodikat seda tüüpi anomaaliate tuvastamiseks. Töö teise osa eesmärk on tuvastada korduvaid radioloogilisi uuringuid. Töös on kasutatud statistikatarkvara R.

# 1 Metoodika

## 1.1 Binoomtest

Antud töös tahetakse näha, kas üldises teenusekasutuses esineb anomaaliaid. Selleks soovitakse luua mudel, mis kirjeldaks üldist teenusekasutust. Selleks jagatakse aasta neljaks perioodiks ning iga perioodi jaoks leitakse sündmuse toimumise tõenäosus. Antud tõenäosus on sellesse perioodi kuuluvate nädalate keskmiste suhteliste sageduste keskmine. Anomaaliate tuvastamiseks kasutatakse binoomtesti, mille  $p$ -väärtuste leidmiseks kasutatakse järgnevat valemit raamatust (Siegel 1956):

$$Pr = \sum_i^N \binom{N}{i} p^i (1-p)^{N-i} \quad (1.1)$$

## 1.2 Keskmine absoluutne viga

Kahe tõenäosusjaotuse erinevuse kirjeldamiseks kasutatakse keskmist absoluutset viga (Cort J. Willmott 2005). Keskmine absoluutne viga leitakse alloleva valemi abil:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - z_i|, \quad (1.2)$$

kus  $n$  on tunnuste arv hulgas,  $y_i$  on hulga  $i$ -ndas vaatlus ja  $z_i$  on hulga  $i$ -ndas vaatlus.

## 1.3 Hierarhiline klasterdamine

Antud peatükk põhineb raamatul (Gareth James 2013). Hierarhilist klasterdamist kasutatakse andmemahtude vähendamiseks ning sarnaselt käituvate gruppide ehk klastrite leidmiseks.

Hierarhilisel klasterdamisel on kaks lähenemist: alt üles ehk aglomeratiivne ja ülevalt alla ehk eristav. Antud töös kasutatakse aglomeratiivset hierarhilist klasterdamist.

Meetodi kirjeldus:

- Iga objekt on eraldi klaster,
- kaks omavahel kõige sarnasemat klastrit ühendatakse,

- omavahel ühendatakse kahte kõige sarnasemat klastrit kuni alles jääb üks suur klaster, kuhu kuuluvad kõik objektid,
- leitud klastrid kujutatakse joonisel dendogrammi või puuna.

Klastrid ühendatakse omavahel nende kauguseid vaadates. Antud töös kasutatakse täieliku seose kaugust, kus klastrite vahelise kauguse leidmiseks leitakse suurim kaugus nende klastrite objektide vahel.

## 2 Globaalse teenusekasutuse anomaaliate otsing

### 2.1 Andmestiku ülevaade

Andmestik on koostatud Eesti Haigekassa 2010 – 2017 aasta raviarvetest, millest on valitud tunnused teenuse kood, teenuse nimetus ning teenuse toimumise kuupäev. Andmed on pseudonümiseeritud nii, et neid ei saa isikutega kokku viia. Andmestikust valitakse ambulatoorsed vastuvõtud, anesteesia, operatsioonid ning uuringud ja protseduurid. Andmestik on viidud kujule, kus igale teenusele on leitud ühe nädala jooksul toimunud antud teenuste arv aastatel 2010 – 2017. Kokku on andmestikus 63 347 rida ja 4 tunnust. Need tunnused on teenuse kood, teenuse nimi, toimumise nädal ja uuringute arv teenuse kohta antud nädalal.

Tabel 2.1: Näidisrida andmestikust koos selgitustega

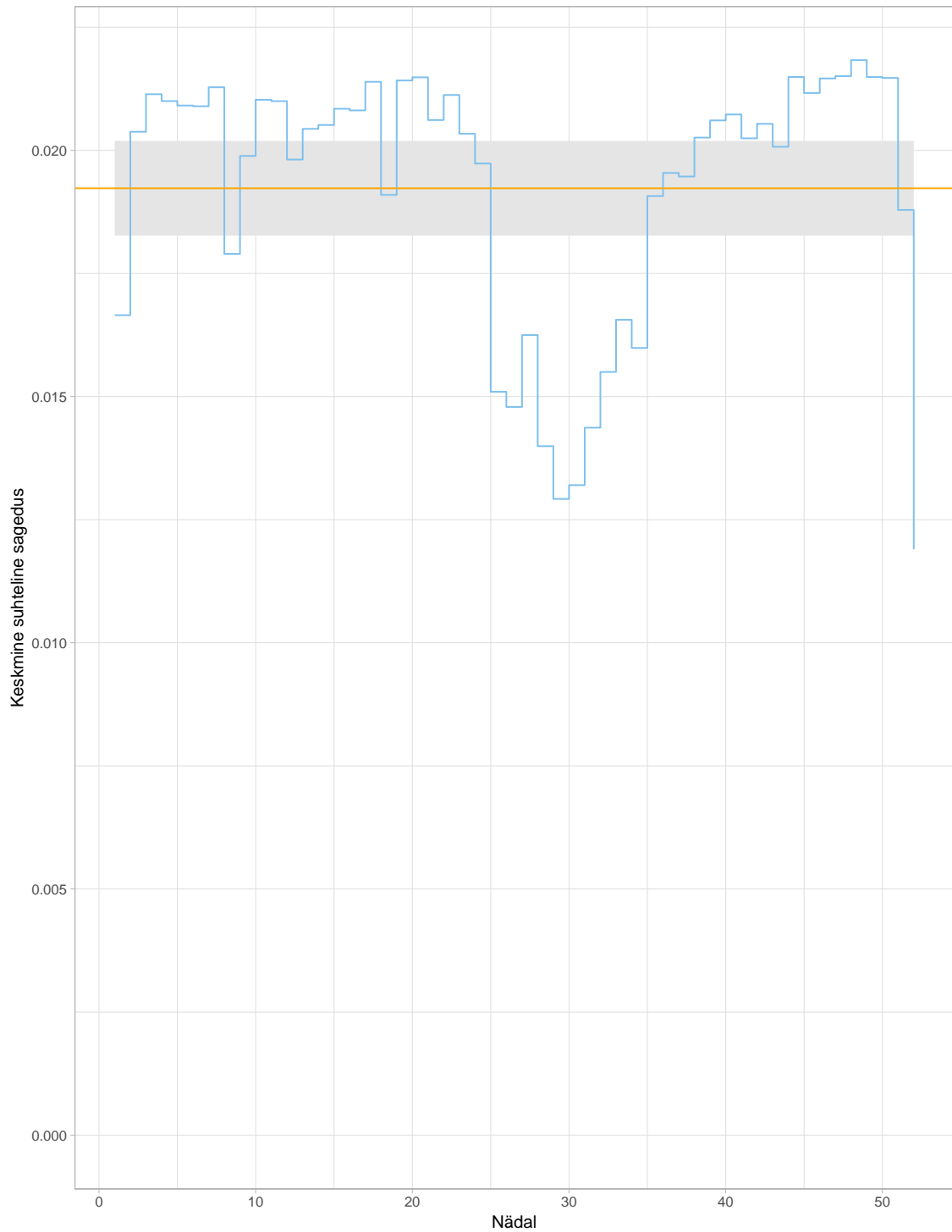
000000000000021002	Teenuse kood
Adenoidektoomia	Teenuse nimetus
1	Nädal
530	Teenuste arv nädalas

Soovitakse näha, kas protseduuride arv nädalas on konstantse sagedusega üle terve aasta. Selleks kasutatakse kirjeldavat analüüsi, kus saadakse ülevaade andmetest.

### 2.2 Üldmudeli fikseerimine

Antud töös uuritakse, kas radioloogilisi uuringuid tehakse aasta lõikes sama sagedusega kui teisi uuringuid. Selle tuvastamiseks on vaja mudelit, mis vastaks üldisele teenusekasutusele.

Eeldus on, et protseduuride arv on aasta lõikes sama sagedusega, ehk igal nädalal tehakse sama palju protseduure. Tahetakse näha, kas see mudel võiks vastata üldisele teenusekasutusele.



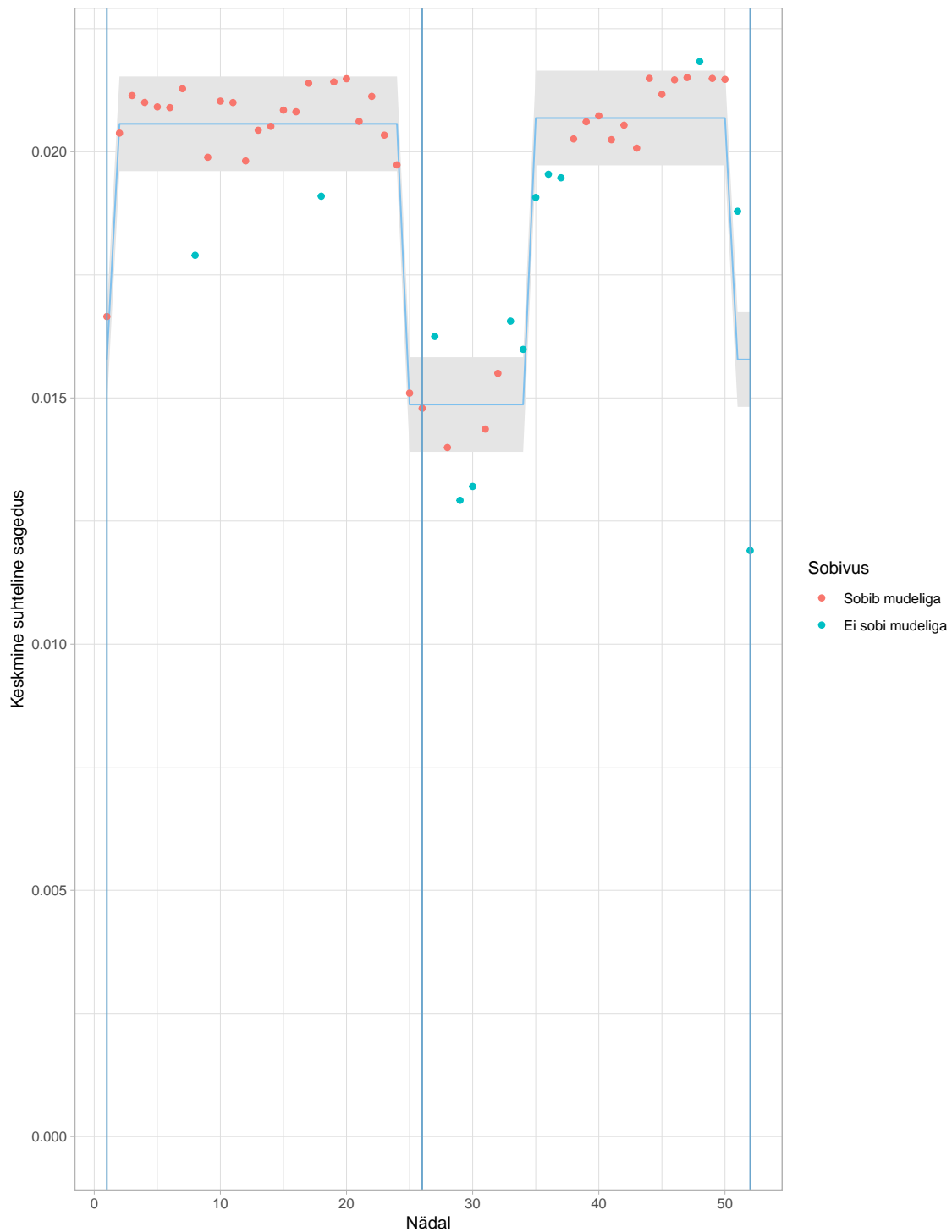
Joonis 2.1: Osutatud teenuste suhteline sagedus nädalate kaupa

Joonisel 2.1 olev oranž joon näitab, milline peaks olema teenusekasutus, kui uuringu-

id tehakse igal nädalal võrdselt. Kindlaks määratud sagedusel  $\frac{1}{52}$  lubatakse varieeruda suuruse 0.00096 võrra, mis määrab lubatus usaldusvahemiku. Antud suurus on 5% eeldatavast sagedusest  $\frac{1}{52}$ . Joonisele 2.1 lisatud usaldusvahemik on (0, 0183; 0, 0202). Kui kõik vaatlused langeksid vastavasse vahemikku, siis oleks eeldus tõene ning saaks öelda, et uuringuid tehakse ühtlase sagedusega. Jooniselt 2.1 on aga näha, et nii ei ole. On selgesti näha, et teenusekasutus ei ole sama sagedusega ning tekivad ajalised mustrid. Üldist teenusekasutust ei saa seetõttu kirjeldada mudeliga, kus igal nädalal toimunud uuringute arv on võrdne.

Eelnevalt on näha, et teenusekasutus on sesoonne ning autor soovib näha, kas aasta vahemikeks jaotamine sobiks üldist teenusekasutust kirjeldavaks mudeliks. Nädalatel 2 – 24 on näha suuremat teenusekasutust, nädalatel 25 – 34 ehk suvel see langeb ning nädalatel 35 – 50 see taas tõuseb. Aasta viimastel nädalatel ning aasta esimesel nädalal on teenusekasutus kõige madalam.

Aasta jaotatakse neljaks osaks eelnevalt defineeritud vahemike järgi ning iga perioodi jaoks arvutatakse selles vahemikus toimunud uuringute keskmine arv. Loodud mudelis viiakse iga nädala kohta läbi modifitseeritud binoomtest, et näha, kas antud mudel vastab tegelikule teenusekasutusele. Antud töös on väga palju observatsioone ning peatükis 1.1 kirjeldatud binoomtesti meetodit tuleb pisut parandada. Etteantavat tõenäosust ei teata ning seetõttu antakse ette intervall, kuhu tõenäosus väärtus sattuma peab. Kui leitud suhteline sagedus  $\hat{p}$  kuulub etteantud intervalli, siis seda vaatlust statistiliselt oluliseks ei peeta. Kui vaatluse suhteline sagedus on etteantud tõenäosuse intervallis olevatest suurustest väiksem, siis tehakse binoomtest intervalli alumise otspunktiga. Kui vaatluse suhteline sagedus on tõenäosuse intervalli ülemiseest piirist suurem, siis valitakse binoomtesti tegemiseks intervalli ülemine otspunkt.



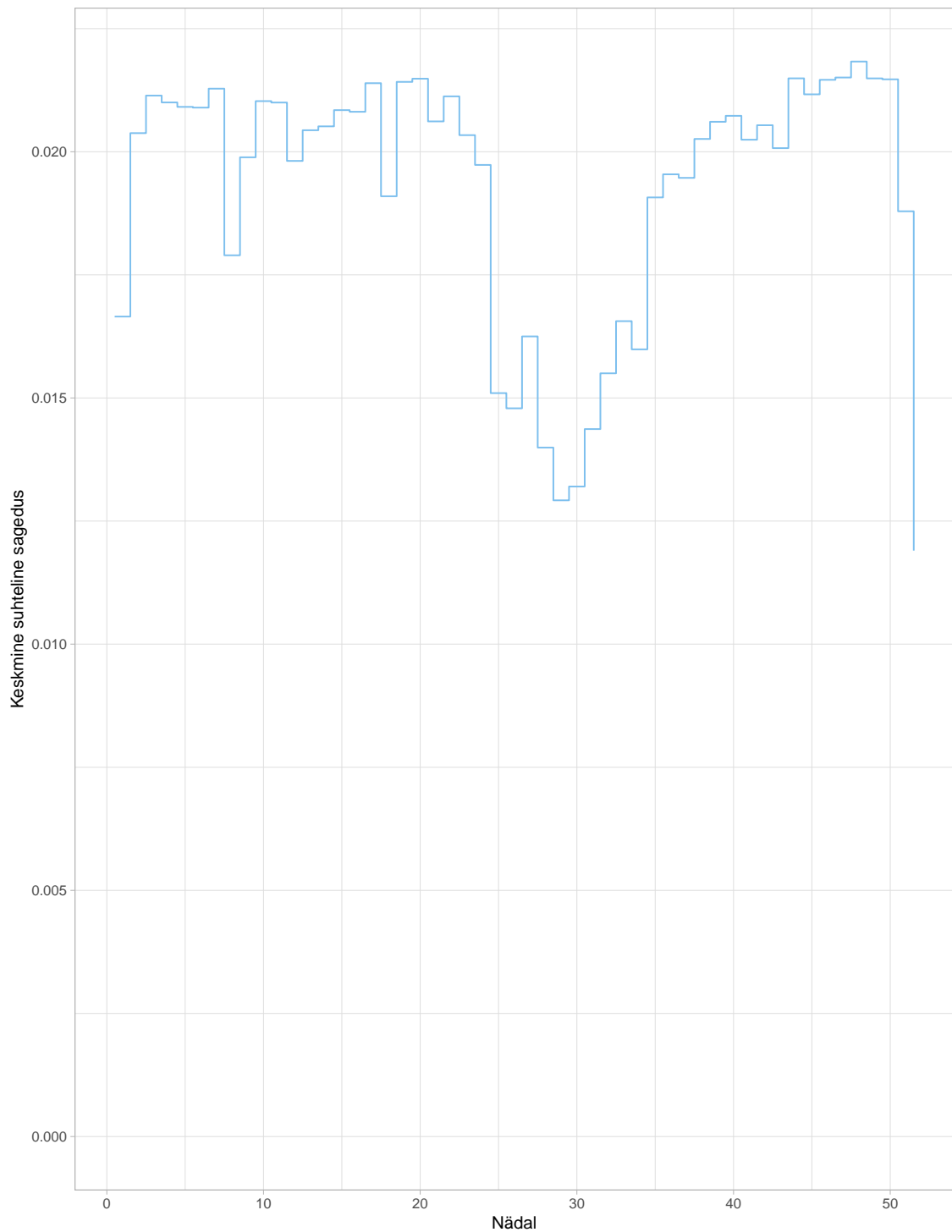
Joonis 2.2: Keskmine osutatud teenuste suhteline sagedus vahemike kaupa

Joonisel 2.2 kujutatud kolm sinist joont on nädalatel 1, 26 ja 52 ning märgivad vastavalt arveldusperioodi algust, esimese arveldusperioodi lõppu ning teise arveldusperioodi lõppu. On näha, et nädal peale esimese arveldusperioodi lõppu tõuseb teenusekasutuse sagedus. Perioodi keskmine suhteline sagedus  $\hat{p}$  võib varieeruda eel-

nevalt defineeritud suuruse 0,00096 võrra. Joonisel 2.2 on halli alana näha nelja keskmise suhtelise sageduse absoluutset vahemikku:

- nädalatel 1, 51 ja 52 on see  $(0,0148; 0,0167)$ ,
- nädalatel 2 – 24 on see  $(0,0196; 0,0215)$ ,
- nädalatel 25 – 34 on see  $(0,0139; 0,0158)$ ,
- nädalatel 35 – 50 on see  $(0,0197; 0,0216)$ .

Siiski on näha, et antud mudel ei kirjelda täielikult üldist teenusekasutust. Selleks, et oleks võimalik näha, kas radioloogilised uuringud vastavad tavapärasele teenusekasutusele on vaja mudelit, mille kõik nädalad on olulised. Selleks leitakse iga nädala suhteline sagedus.



Joonis 2.3: Osutatud teenuste suhteline sagedus

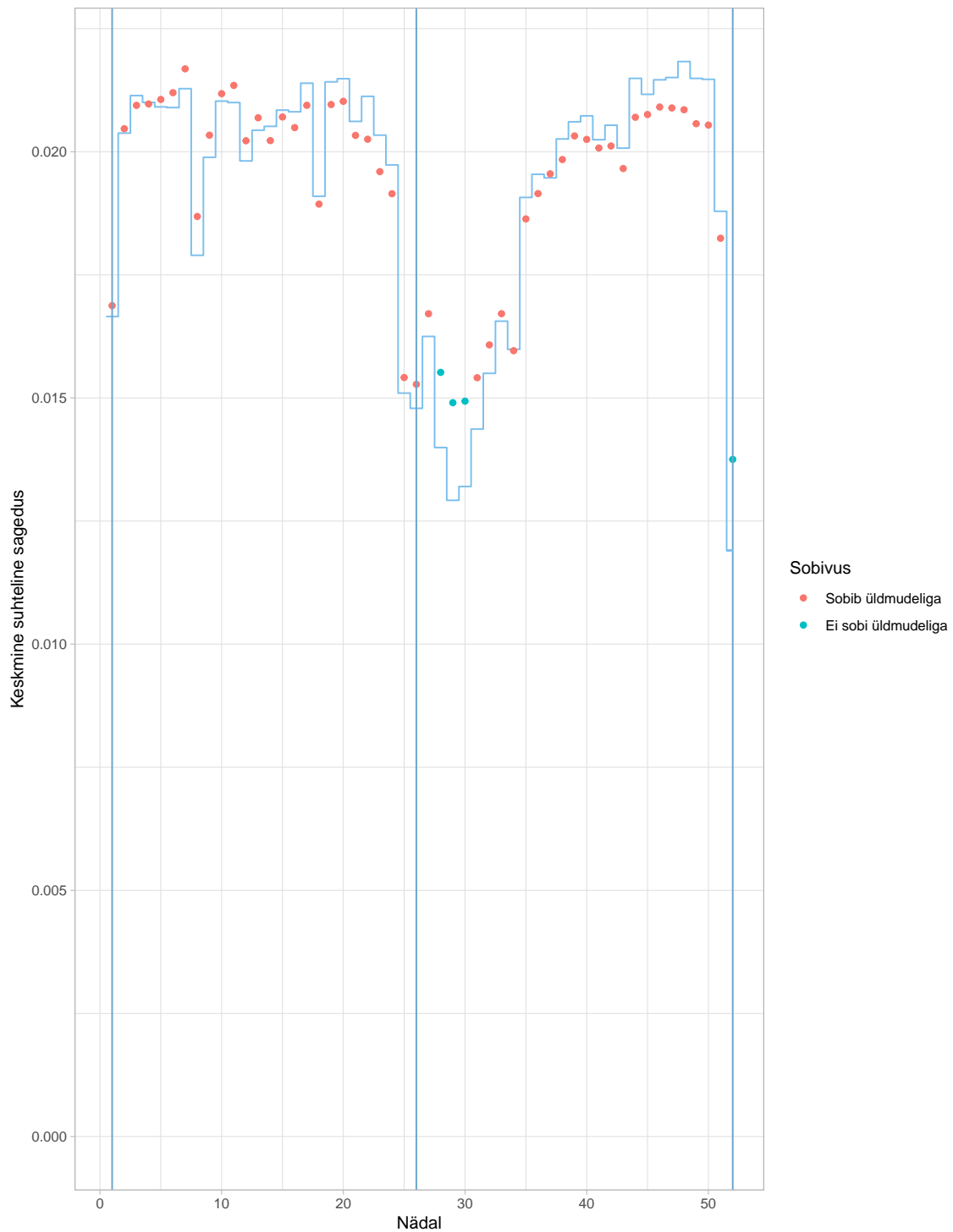
Kuna joonis 2.3 vastab üldisele teenusekasutusele, siis võrreldakse üldist radioloogiliste uuringute tegemist just sellega.

## 2.3 Radioloogilise teenusekasutuse võrdlus globaalse teenusekasutusega

Antud töös keskendutakse radioloogilistele uuringutele ning tahetakse teada, kas radioloogilised uuringud vastavad tavapärasele teenusekasutuse mudelile. Radioloogilistest uuringutest vaadeldakse antud töös röntgenit, ultraheli, kompuutertomograafiat ja magnetresonantstomograafiat. Algsest andmestikust eraldati radioloogilisi teenuseid sisaldavad read. Radioloogiliste teenuste andmestikus on 36 934 rida ja 6 tunnust. Need tunnused on teenuse kood, teenuse nimetus, teenuse toimumise nädal, teenuse toimumise aasta, ühe nädala jooksul tehtud uuringute arv ning radioloogia tüüp. Edasi töötatakse just nende eraldatud ridadega ja vaadeldakse iga aasta radioloogiliste uuringute arvu nädalas, mida võrreldakse üldmudeliga. Selleks, et otsustada, kas vaadeldav nädal vastab üldisele teenusekasutusele, kasutatakse eelnevalt kirjeldatud modifitseeritud binoomtesti. Iga nädala korral on binoomtesti tegemiseks erinev tõenäosuste vahemik, mille keskpunkt on antud nädala üldmudeli suhteline sagedus. Taaskord võib suhteline sagedus varieeruda suuruse 0,00096 võrra.

Tabel 2.2: Näidisrida radioloogiliste uuringute andmestikust koos selgitustega

0000000000000007900	Teenuse kood
Röntgeniülesvõte peapiirkonnast (üks ülesvõte)	Teenuse nimetus
1	Nädal
2016	Aasta
1291	Teenuste arv nädalas
Röntgen	Radioloogia tüüp

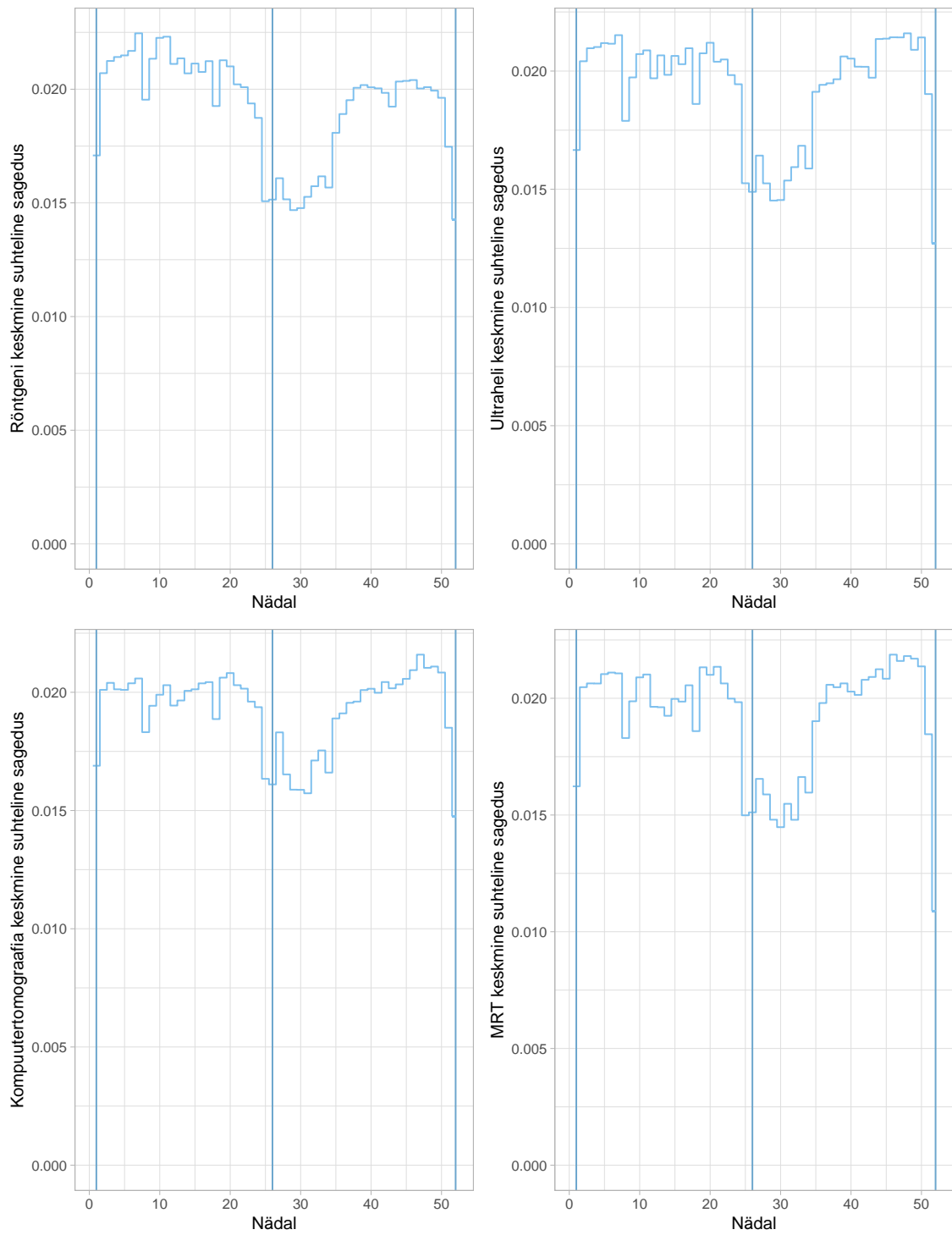


Joonis 2.4: Radioloogiliste uuringute sobivus üldmudeliga

Kuigi radioloogilised uuringud käituvad üldmudeliga üsna sarnaselt, siis jooniselt 2.4 on näha, et nädalatel 28 – 30 ehk juulis ning aasta viimasel nädalal see nii pole.

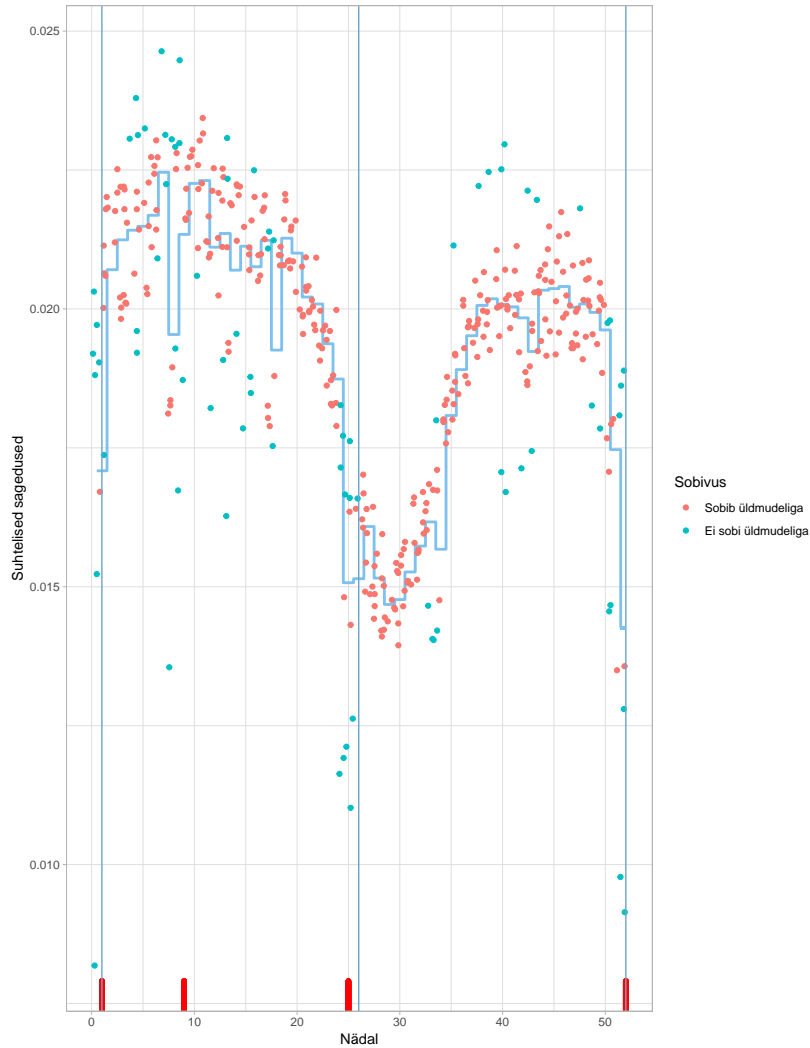
## 2.4 Radioloogiliste uuringute teenusekasutus aastate lõikes

Edasi uuritakse radioloogiliste uuringute teenusekasutuse muutust aastatel 2010 – 2017, et tuvastada ajalisi anomaaliaid. Kaheksa aasta keskmist teenusekasutust vaadates muutuvad ühes aastas toimunud ebastandardised sündmused mitteoluliseks ning ei mõjuta üldist tulemust. Mitme aasta keskmist vaadates saab paremini aru, kuidas radioloogilised teenused üldiselt käituda võiksid. Selleks luuakse iga tüübi kohta eraldi üldmudel 2010 – 2017 aasta keskmise suhtelise sageduse põhjal, mida on näha joonisel 2.5.



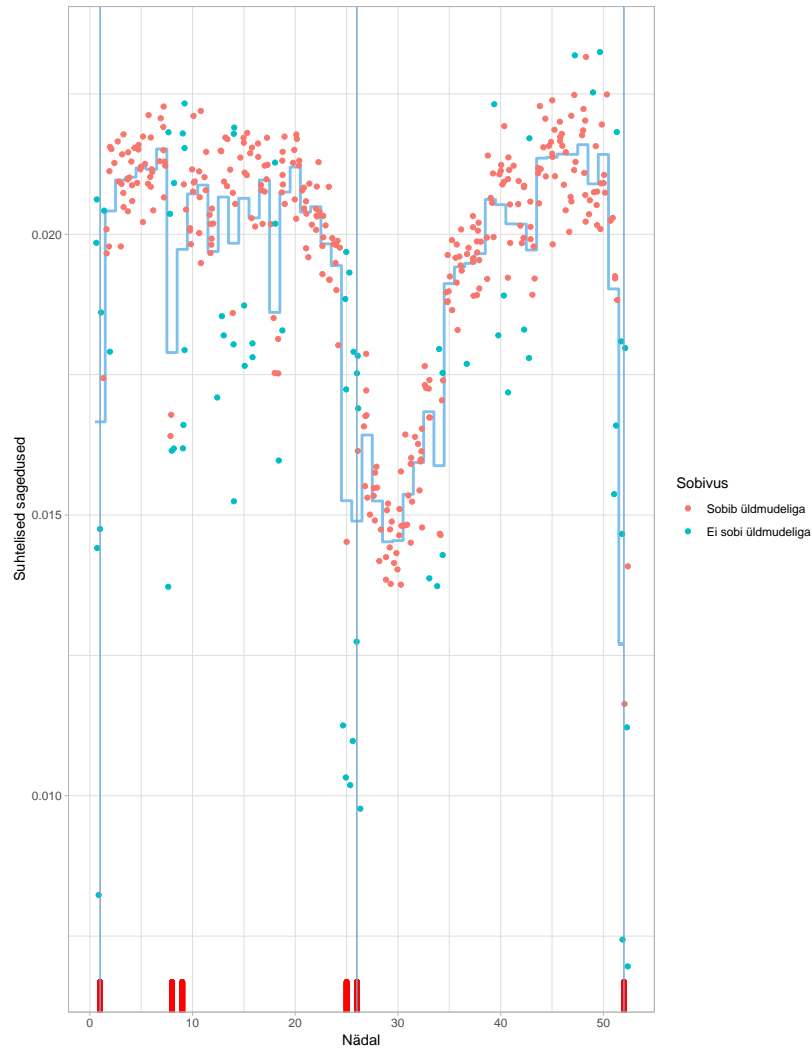
Joonis 2.5: Radioloogiliste uuringute keskmised suhtelised sagedused

Aastate 2010 – 2017 igal nädalal tehtud radioloogiliste uuringutele (eraldi röntgen, ultraheli, kompuutertomograafia ja magnetresonantstomograafia) viiakse läbi eelnevalt selgitatud modifitseeritud binoomtest, kus tõenäosuste intervallid leitakse nii nagu eelmises peatükis. Modifitseeritud binoomtest tehakse iga aasta kohta eraldi. Saadud tulemused kujutatakse graafiliselt kaheksa aasta lõikes.



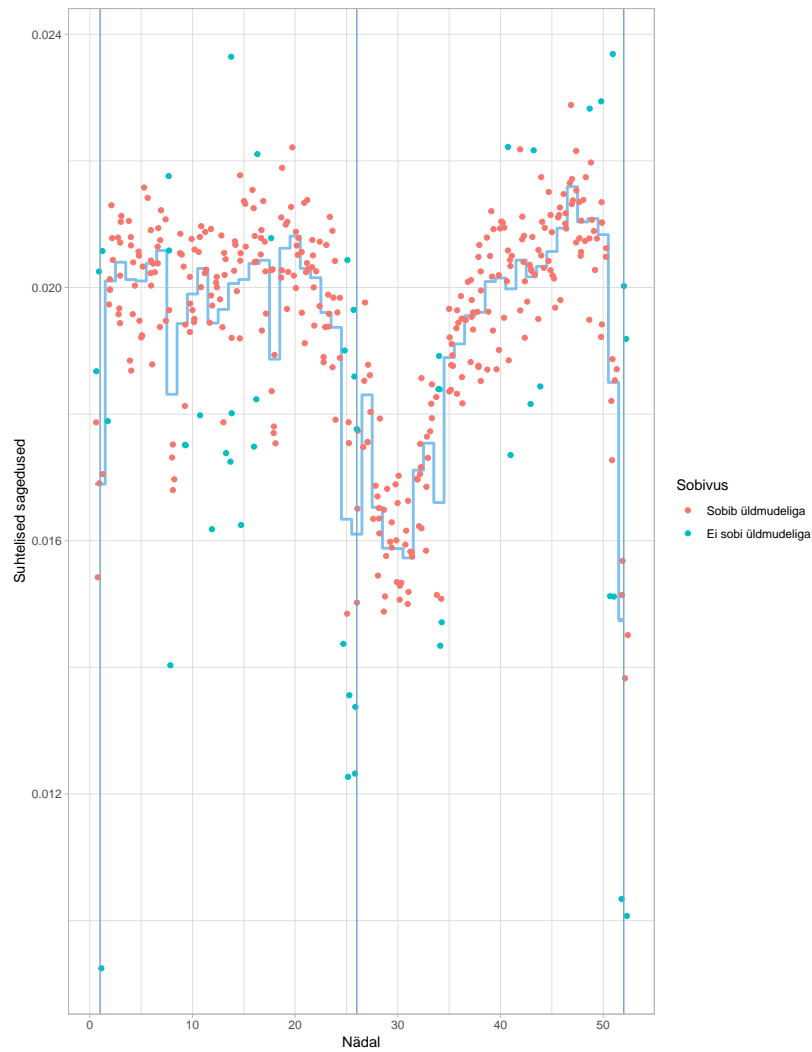
Joonis 2.6: Röntgenuuringu sobivus aastatel 2010-2017 röntgeni üldmudeliga

Esmalt vaadeldakse röntgenuuringuid. Jooniselt 2.6 on näha, et röntgenuuringu tegemine ei järgi üldist mudelit ja ei ole aastate lõikes ühtlane. Joonise x-teljel olevad punased jooned tähistavad nädalaid, kus nullmudelist erinevaid vaatlusi on kuus või enam. See näitab, millistel nädalatel varieerub teenusekasutus kõige rohkem. Röntgenuuringu puhul on kõige rohkem varieeruvad nädalad 1, 8, 25 ja 52.



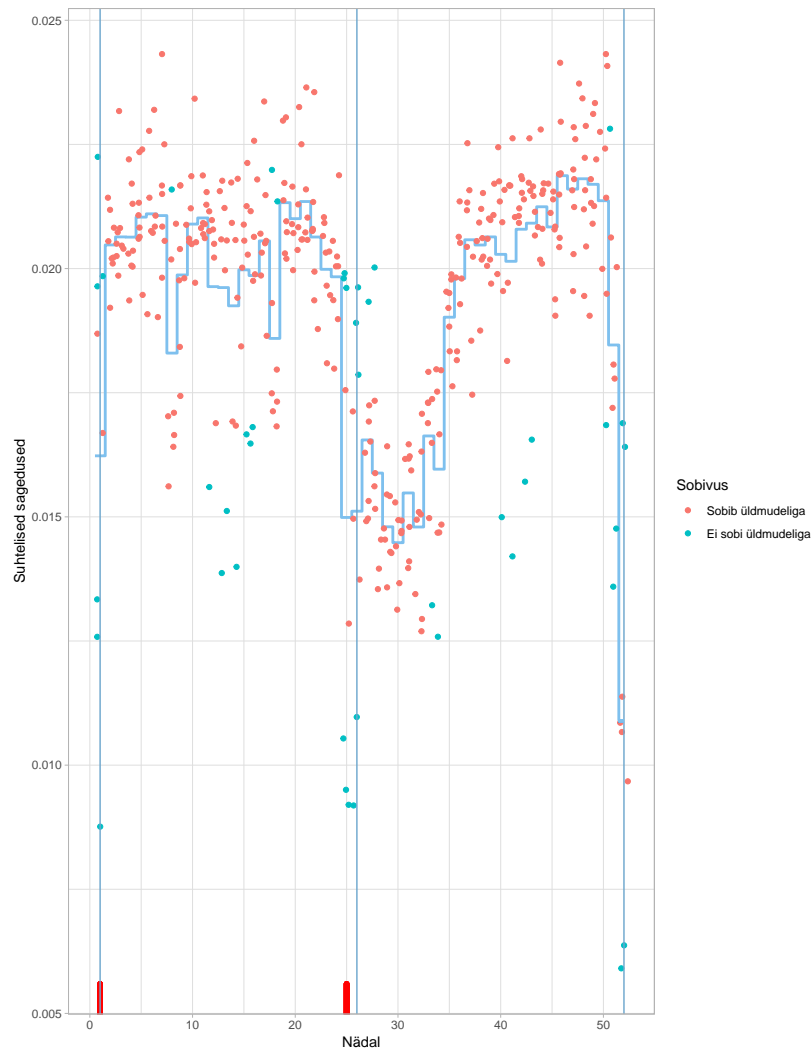
Joonis 2.7: Ultraheliuuringute sobivus aastatel 2010-2017 ultraheli üldmudeliga

Järgmisena vaadeldakse ultraheliuuringute varieeruvust joonisel 2.7. On näha, et ka ultraheliuuringute teenusekasutus varieerub aastate lõikes ning suurima erinevusega on nädalad 1, 8, 9, 25, 26 ja 52, mis on vastavalt aasta esimene nädal, Eesti Vabariigi aastapäev, Jaanipäev, esimese arveldusperioodi lõpp ning aasta viimane nädal. Ultraheliuuringud ei vasta igal aastal üldisele mudelile.



Joonis 2.8: Kompuutertomograafia uuringute sobivus aastatel 2010-2017 kompuutertomograafia üldmudeliga

Jooniselt 2.8 on näha, et iga aasta puhul eraldi ei vasta loodud mudel kompuutertomograafia teenusekasutuse kohta tegelikkusele. Kompuutertomograafia uuringute tegemise varieerus ei ole väga suur, sest ei ole ühtegi nädalat, kus nullmudelile mittevastavate nädalate arv oleks suurem kui kuus.

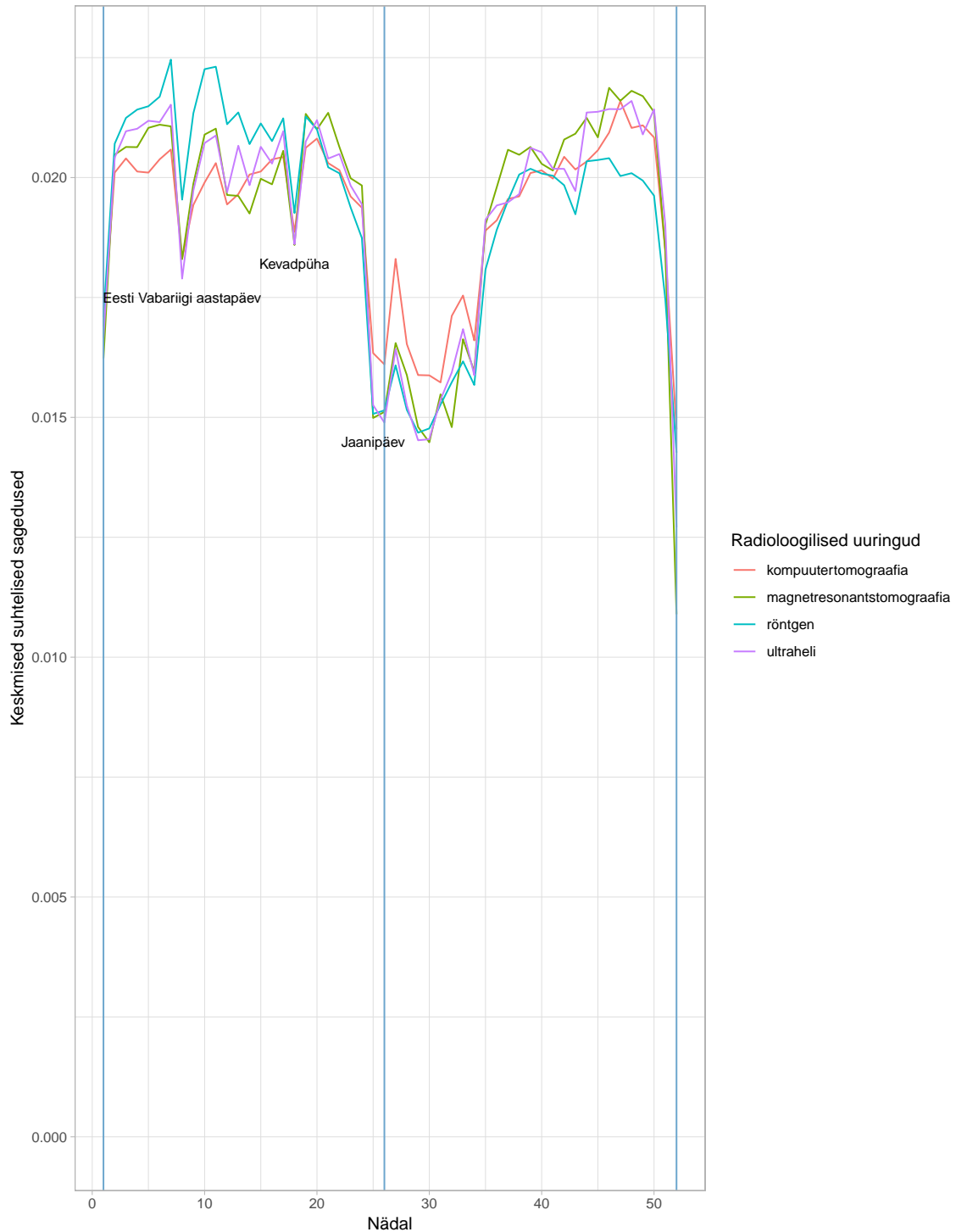


Joonis 2.9: Magnetresonantstomograafia uuringute sobivus aastatel 2010-2017 magnetresonantstomograafia üldmudeliga

Kuigi magnetresonantstomograafia teenusekasutus on aastate lõikes varieeruv, siis on näha, et enamasti on mudelile mittevastavate nädalate suhteline sagedus madalam kui keskmine suhteline sagedus. Nädalad, millal nullmudelile mittevastavaid vaatlusi on rohkem kui kuus, on aasta esimene nädal ning nädal, millal on Jaanipäev.

Radioloogiliste uuringute puhul on näha, et olenevalt aastast, on varieeruvus üsna suur ning kõik aastad ei vasta üldisele teenusekasutusele. Eelnevatelt joonistelt ei ole võimalik näha, mis aastal üldmudelile mittevastavad uuringud toimusid ning seetõttu ei saa öelda, mis aastatel enim üldisest teenusekasutusest erinevaid vaatlusi oli. Eelnevalt on jooniselt 2.4 näha, et radioloogiliste uuringute teenusekasutus on sarnane üldmudelile. Soovitakse teada, kui suur on radioloogiliste gruppide (röntgen,

ultraheli, kompuutertomograafia ja magnetresonantstomograafia) omavaheline varieeruvus. Selleks leitakse iga uuringutüübi aastate 2010 – 2017 keskmine suhteline sagedus.



Joonis 2.10: Radioloogiliste gruppide võrdlus

Joonisel 2.10 kujutatud radioloogiliste teenuste gruppide keskmised suhtelised sa-

gedused järgivad sama mustrit, mida on eriti hästi näha 8. nädalal, kus on Eesti Vabariigi aastapäev, 18. nädalal, kus on Kevadpüha ja 25. nädalal ehk Jaanipäeval. Joonisel 2.10 joonistub ka väga selgelt välja suveperiood, kus inimesed puhkavad. On näha, et kevadel viiakse läbi rohkem röntgenuuringuid, suvel kompuutertomograafia uuringuid ning sügisel tehakse enim magnetresonantstomograafia- ja ultraheliuuringuid.

Selleks, et leida uuringutüüpide mudelite erinevused, leitakse kahe radioloogilise tüübi keskmine absoluutne viga ning maksimaalne suhtelise sageduse vahe. Erinevate radioloogiliste uuringute leitud keskmised suhtelised sagedused on vaadeldavad kui tõenäosuste jaotused. Tõenäosuste jaotuste võrdlemiseks on peatüki 1.2 alusel valitud keskmine absoluutne viga (MAE). Seejärel leitakse, kui suure osa moodustab keskmine absoluutne viga  $\frac{1}{52}$ -st järgneva valemi abil, kus valemis 1.2 suuruse  $n$  väärtus on 52:

$$\text{Keskmise absoluutse vea osakaal} = \frac{MAE}{\frac{1}{52}} \cdot 100\% \quad (2.1)$$

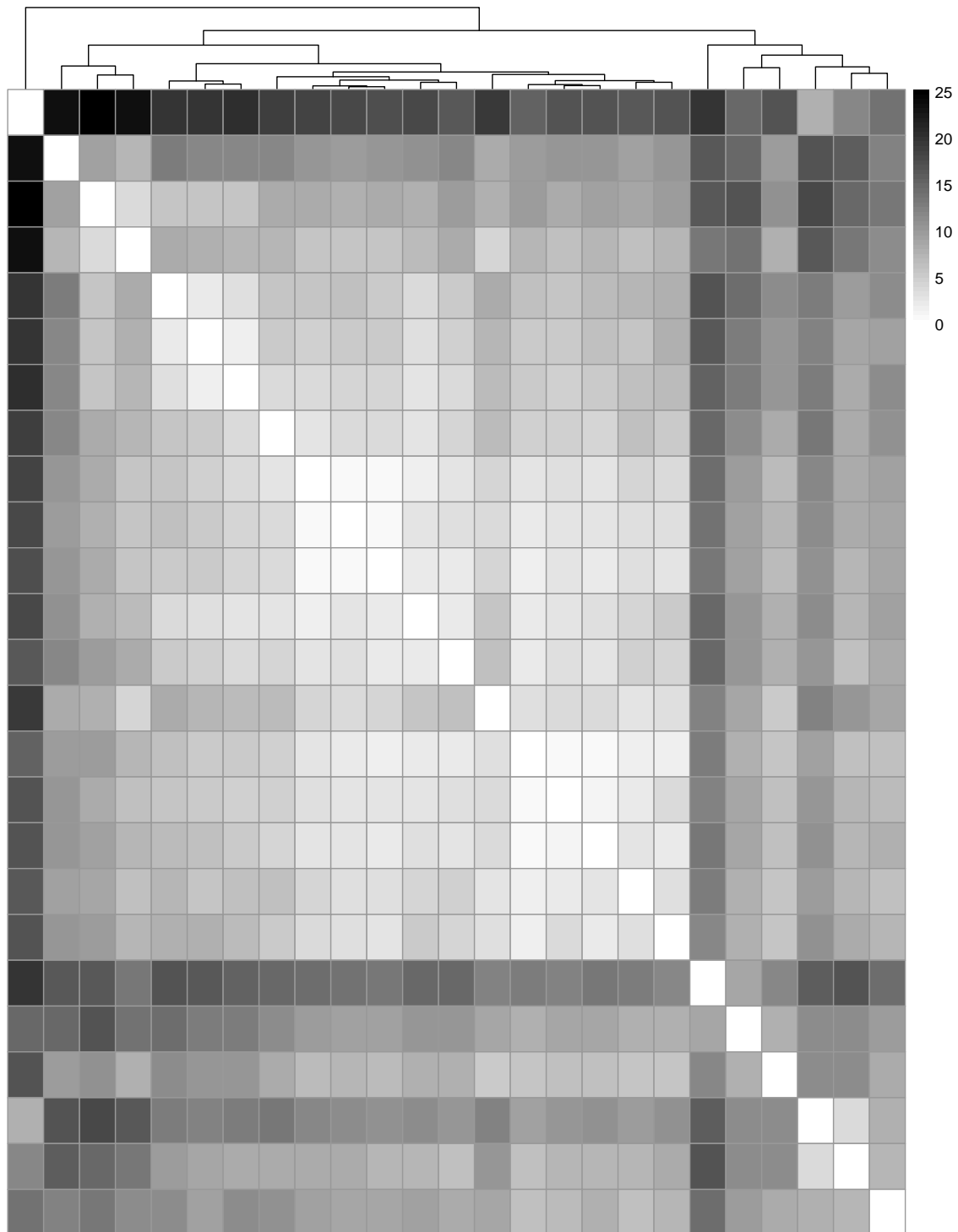
Tabel 2.3: Uuringutüüpide keskmise absoluutse vea ja maksimaalse suhtelise sageduse vahe võrdlus.

Radioloogilise uuringu tüübid	Keskmise absoluutse vea osakaal	Maksimaalne suhtelise sageduse vahe
Röntgen ja ultraheli	3,481%	0,0018
Röntgen ja kompuutertomograafia	4,610%	0,0024
Röntgen ja magnetresonantstomograafia	4,609%	0,0034
Ultraheli ja kompuutertomograafia	3,101%	0,0020
Ultraheli ja magnetresonantstomograafia	2,069%	0,0018
Kompuutertomograafia ja magnetresonantstomograafia	3,543%	0,0039

## 2.5 Radioloogiliste uuringute klasterdamine

Autor soovib teada, kas radioloogilised teenused grupeeruvad keskmise absoluutse vea järgi. Selleks leitakse 25 kõige sagedasemat radioloogilist uuringut, mis moo-

dustavad radioloogiliste uuringute andmestikus olevatest andmetest umbes 83%. Selleks, et vähendada vaatluste varieeruvust, kasutatakse kärbitud keskmist. Kui eelnevalt kasutati kaheksa aasta lõikes leitud keskmist, siis nüüd eemaldatakse keskmise leidmiseks iga nädala minimaalne ja maksimaalne suhteline sagedus kaheksa aasta lõikes. Seejärel leitakse uuesti iga radioloogiatüübi keskmine suhteline sagedus. Seejärel leitakse nende teenuste omavahelised keskmised absoluutsed vead ning nende osakaalu  $\frac{1}{52}$ -st. Saadud tulemused kujutatakse intensiivsuskardil ning nende grupeerimiseks kasutatakse hierarhilist klasterdamist.



Joonis 2.11: 25 sagedasema radioloogilise teenuse keskmise absoluutse vea suhte intensiivsuskart

Jooniselt 2.11 on näha, et tekib kaks selgelt eristuvat klastrit ja üks erind. Esimeses klastris on 6 teenust ja teises klastris 18 teenust.

Esimeses klastris olevad radioloogilised uuringud on:

- silma ultraheliuuring,
- silmapõhja optiline koherentne tomograafia,
- röntgeniülesvõtte rindkere piirkonnast (7903),
- röntgeniülesvõtte rindkere piirkonnast (7904),
- kilpnäärme ultraheliuuring,
- kõhupiirkonna ultraheliuuring.

Teises klastris olevad radioloogilised uuringud on:

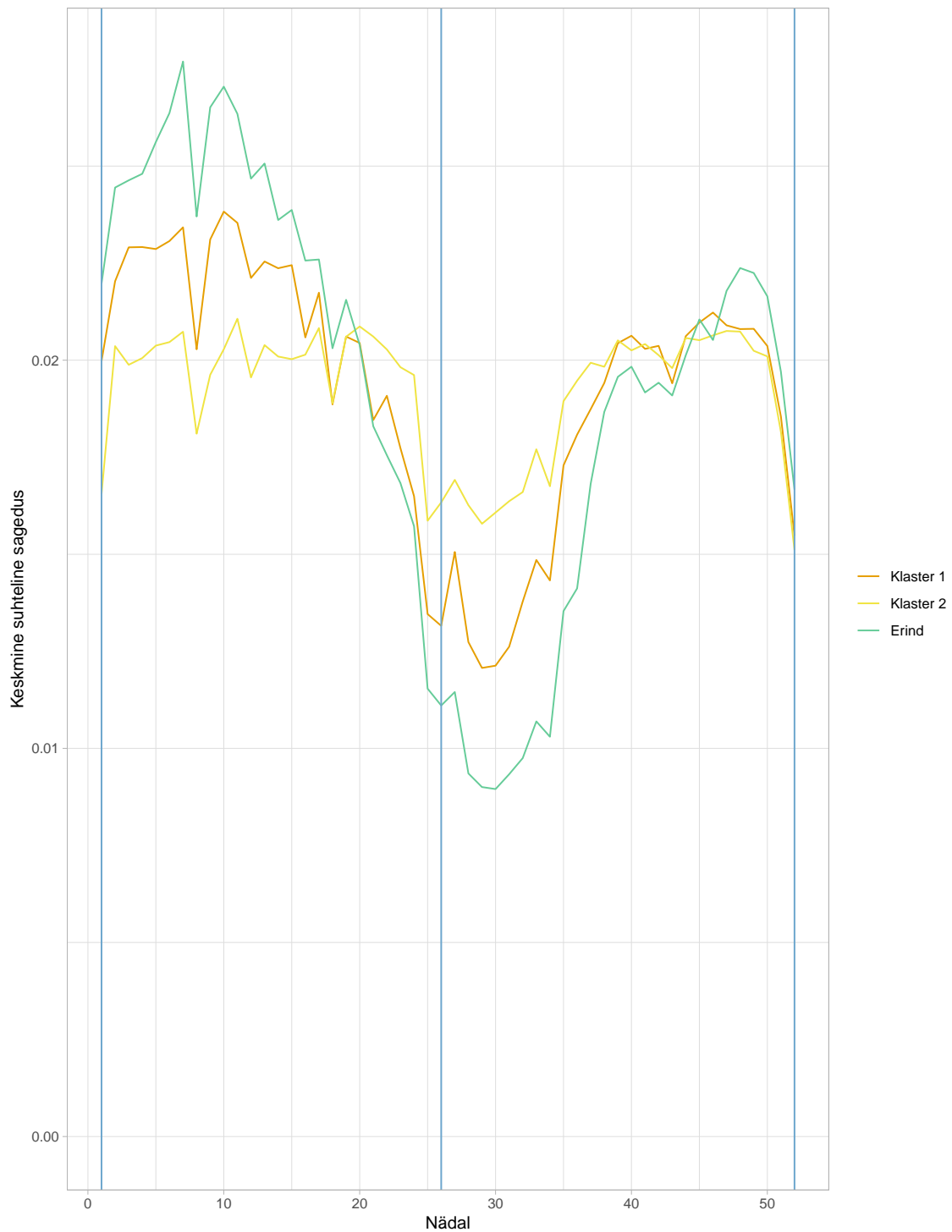
- mammograafia, üks rinnanääre kahes sihis,
- röntgeniülesvõtte lülisamba piirkonnast,
- röntgeniülesvõtte vaagnapiirkonnast,
- röntgeniülesvõtte ülajäsemetest,
- röntgeniülesvõtte alajäsemetest (7919),
- röntgeniülesvõtte alajäsemetest (7920),
- ühe piirkonna arterite ultraheliuuring,
- raseduse ultraheliuuring,
- pehmete kudede ultraheliuuring,
- kõhu- ja vaagnapiirkonna ultraheliuuring,
- vaginaalne ultraheliuuring,
- kompuutertomograafia natiivis (7975),
- kompuutertomograafia natiivis (7976),
- kompuutertomograafia kontrastainega (7978),
- kompuutertomograafia kontrastainega (7979),
- kompuutertomograafia-angiograafia,
- peaaju kompuutertomograafia natiivis,
- kompuutertomograafia kontrastaine 10 ml.

Uuring, mis on teistest eristuv, on:

- röntgeniülesvõtte peapiirkonnast.

Autor arvas, et sama tüüpi radioloogilised uuringud võiksid käituda sarnaselt, kuid on näha, et nii ei ole. Ühte klastrisse kuuluvad röntgen-, kompuutertomograafia- ja

ultraheliuuringud. Teine ootus oli, et erinevatele kehaosadele tehtavad radioloogilised uuringud võiksid aasta lõikes sarnase sagedusega olla. Klasterdamise tulemusena on näha, et kuigi kõik kehaosad ei grupeeru eraldi klastritesse, siis teatud muster esineb. Näiteks esimesse klastrisse kuuluvat rindkere ülesvõtet ei esine teistes klastrites. Samuti kuuluvad ala- ja ülajäsemetele tehtud uuringud vaid teise klastrisse.

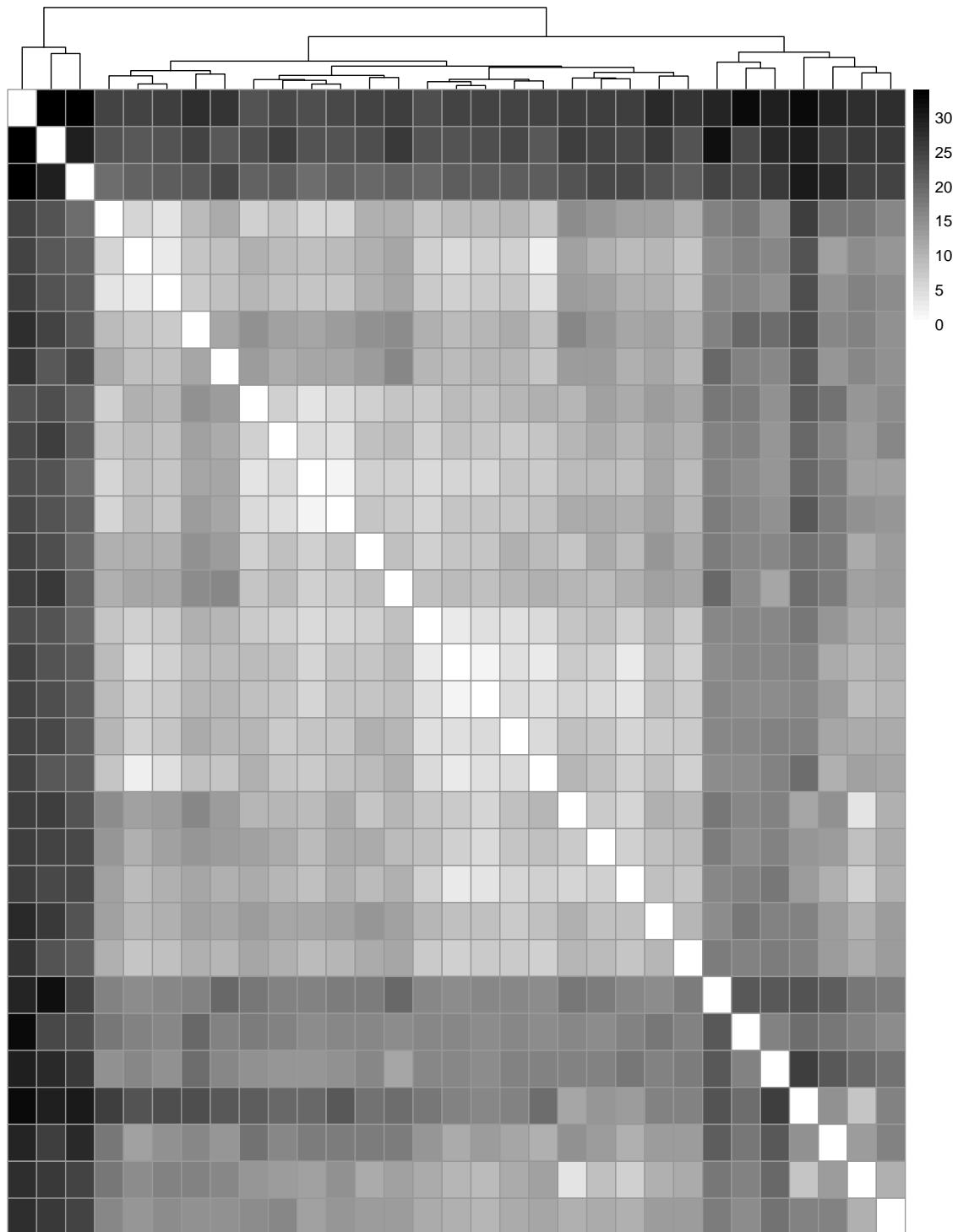


Joonis 2.12: Kahe klastri ja ühe erindi keskmise sageduse võrdlus aastate 2010-2017 lõikes

Jooniselt 2.12 on näha leitud kolme klastri keskmise suhtelise sageduse erinevust aastate 2010 – 2017 lõikes. Eelnevalt leiti iga uuringu iga aasta nädala minimaalne ja maksimaalne suhteline sagedus kaheksa aasta lõikes, mis seejärel suhtelise keskmise

arvutamisest välja arvati. Alles jäänud andmete põhjal leiti joonisel 2.12 kujutatud keskmised suhtelised sagedused. On näha, et üldine muster on klastrite lõikes sarnane, kuid sesoonne varieeruvus klastrites erineb. Kõige suurema amplituudiga on kolmas klaster ning vähima amplituudiga on teine klaster.

Autoril ei õnnestunud leida, et radioloogiliste uuringute klasterdamisel jaguneksid erinevad tüübid erinevatesse klastritesse. Seetõttu soovitakse vaadata ühte radioloogia tüüpi lähemalt. Selleks valitakse röntgen, sest see on kõige sagedasem uuring. Uuringute kohta leitakse aastatel 2010 – 2017 iga nädala suhteline sagedus ja eemaldatakse maksimaalne ja minimaalne vaatlus, et vähendada varieeruvust. Seejärel leitakse üle kõikide röntgenuuringute keskmised suhtelised vead ja kui suure osa moodustab keskmine absoluutne viga  $\frac{1}{52}$ -st ning saadud suurused kujutatakse intensiivsuskardil. Kokku on röntgenuuringuid, kus kaheksa aasta jooksul tehakse igal nädalal vähemalt kaks uuringut, 31 tükki. Klasterdamise abil leitakse ka sarnaselt käituvad grupid.



Joonis 2.13: Röntgenuuringute teenuse keskmise absoluutse vea suhte intensiivsuskart

Hierarhilise klasterdamise tulemusena on jooniselt 2.13 näha, et tekib neli eristuvat klastrit. Esimesse klastrisse kuulub 21 uuringut, teise klastrisse 7 uuringut ja kolmandasse klastrisse 2 uuringut. Lisaks on üks erinduuring, mis eelnevatesse klaster-

ritesse ei sobitu. Esimesse klastrisse kuuluvad uuringud on:

- mammograafia, üks rinnanääre kahes sihis,
- röntgeniülesvõte peapiirkonnast (7901),
- röntgeniülesvõte rindkere piirkonnast (7904),
- röntgeniülesvõte lülisamba piirkonnast (7906),
- röntgeniülesvõte lülisamba piirkonnast (7907),
- röntgeniülesvõte lülisamba piirkonnast (7908),
- röntgeniülesvõte kõhupiirkonnast (7909),
- röntgeniülesvõte vaagnapiirkonnast (7912),
- röntgeniülesvõte vaagnapiirkonnast (7913),
- röntgeniülesvõte vaagnapiirkonnast (7914),
- röntgeniülesvõte ülajäsemetest ja/või liigestest (7915),
- röntgeniülesvõte ülajäsemetest (7916),
- röntgeniülesvõte ülajäsemetest (7917),
- röntgeniülesvõte alajäsemetest (7918),
- röntgeniülesvõte alajäsemetest (7919),
- röntgeniülesvõte alajäsemetest (7920),
- mittetransporditava haige röntgeniülesvõte väljaspool röntgenikabinetti (7923),
- mittetransporditava haige röntgeniülesvõte väljaspool röntgenikabinetti täisdigitaalsel seadmel (7925),
- röntgenoskoopia (üks piirkond),
- söögitoru ja mao või peensoole kaksikkontrasteerimine röntgenoskoopial,
- röntgenoskoopia kontrastainega, sh fistulograafia.

Teise klastrisse kuuluvad järgmised röntgenuuringud:

- röntgeniülesvõte peapiirkonnast (7900),
- röntgeniülesvõte peapiirkonnast (7902),

- röntgeniülesvõte rindkere piirkonnast (7903),
- röntgeniülesvõte rindkere piirkonnast (7905),
- röntgeniülesvõte kõhupiirkonnast (7910),
- mittetransporditava haige iga järgnev röntgeniülesvõte väljaspool röntgenikabinetti,
- jäseme ülesvõte telje mõõtmiseks või täispikkuses lüüsisamba röntgeniülesvõte.

Kolmandas klastris on röntgenuurid:

- röntgeniülesvõte kõhupiirkonnast (7911),
- vastsündinu röntgeniülesvõte.

Teistest eristuv röntgenuurid on:

- mittetransporditava haige iga järgnev röntgeniülesvõte väljaspool röntgenikabinetti täisdigitaalsel seadmel.

Autor arvas, et kui jagada uurid klastritesse radioloogiatüübi alusel, siis eristuvad tehtud uurid kehasade järgi paremini. Saadud tulemused aga seda ei näita.

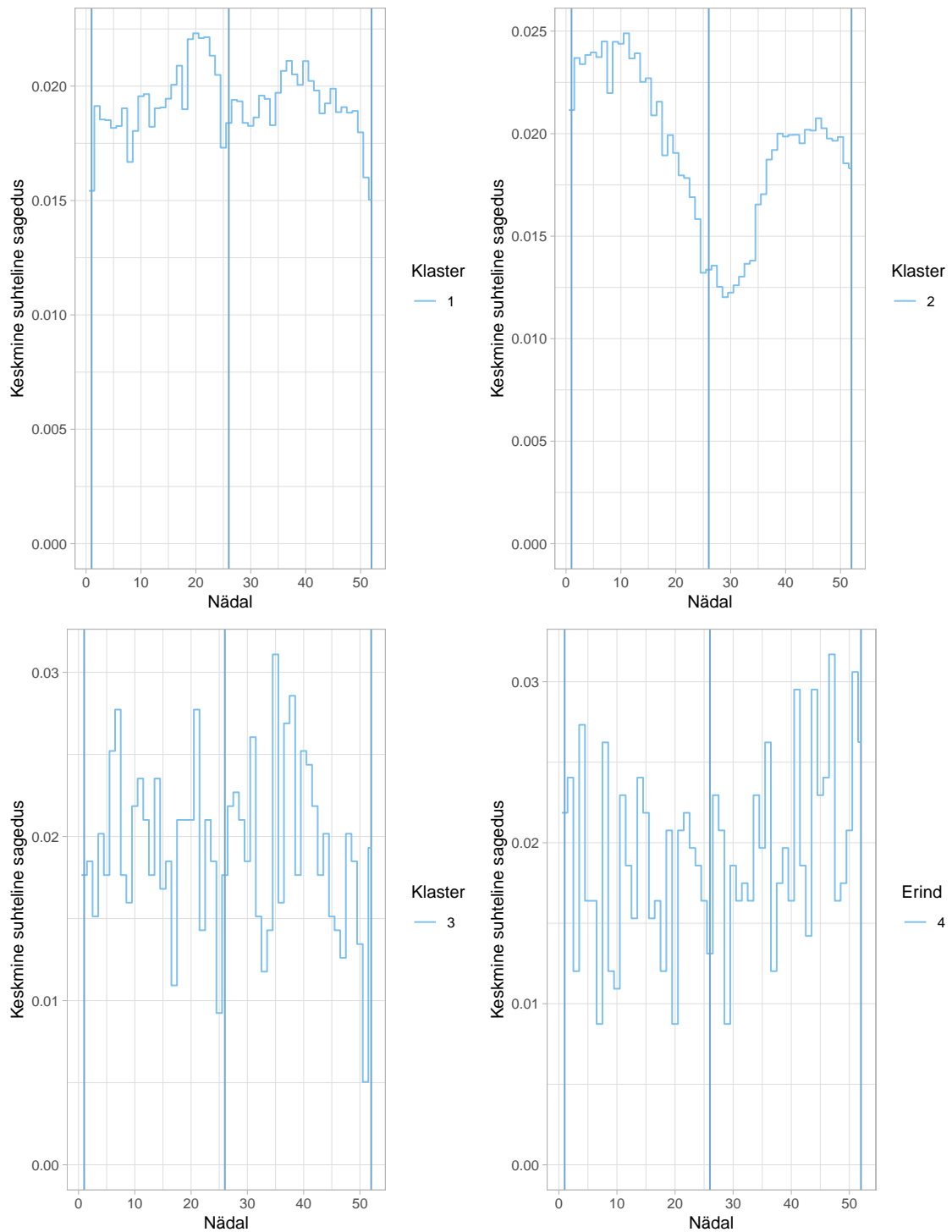


Joonis 2.14: Röntgenuringute kolme klasteri keskmise sageduse võrdlus aastate 2010-2017 lõikes

Joonisel 2.14 on kujutatud röntgenuringute kolme klasteri keskmised suhtelised sagedused, mis on leitud kärbitud keskmiste suhteliste sageduste kaudu. On näha, et esimene klaster varieerub kõige vähem ja teine klaster on kõige sesoonsem. Erin-

duuringu keskmist suhtelist sagedust ei lisatud joonisele 2.14, et klastrite erinevused paremini välja tuua. Jooniselt 2.14 on näha, et kolmandal klastril pole kindlat se-  
soonset trendi. Võib järeldada, et uuringuid tehakse vastavalt vajadusele ning need  
ei ole edasilükatavad.

Soovitakse vaadata iga klatri sisse ning valitakse igast klastrist kõige sagedasem  
uuring. Seejärel leitakse antud uuringu keskmine suhteline sagedus ning saadud tu-  
lemused on kujutatud alloleval joonisel 2.15 koos eristuva uuringu keskmise suhtelise  
sagedusega.



Joonis 2.15: Kolme klastris ühe uuringu ja erinduuringu suhteline sagedus aastate 2010-2017 lõikes

Esimeses klastris on kujutatud uuringu röntgenülesvõtte alajäsemetest (kaks ülesvõtet) keskmine suhteline sagedus, teises klastris uuringu röntgenülesvõtte rindkere piirkonnast (üks ülesvõte) keskmine suhteline sagedus, kolmandas klastris uuringu vastsündinu

röntgeniülesvõte (üks ülesvõte) keskmine suhteline sagedus ning neljandal joonisel erinduuringu mittetransporditava haige iga järgnev röntgeniülesvõte väljaspool röntgenikabinetti täisdigitaalsel seadmel keskmine suhteline sagedus.

## 2.6 Järeldus

Eesmärk oli leida mudel, mille abil oleks võimalik tuvastada ajalisi anomaaliaid. Antud töös loodi kolm mudelit, mis võiks kirjeldada üldist teenusekasutust. Esimene mudel eeldas, et uuringuid tehakse aastas sama sagedusega, kuid see ei kirjeldanud üldist teenusekasutust. Teises mudelis jaotati aasta neljaks vahemikuks ning igasse vahemikku kuuluvatele nädalatele leiti tõenäosus, et uuring tehakse antud nädalal. Tõenäosus leiti vahemiku keskmise suhtelise sageduse järgi. Ka see mudel ei kirjeldanud üldist teenusekasutust. Seejärel leiti iga nädala kaheksa aasta keskmine suhteline sagedus ning see loeti tõenäosuseks, et uuring tehakse antud nädalal. Võrreldes radioloogilisi uuringuid üldmudeliga leiti, et radioloogilised uuringud ei järgi üldist teenusekasutust.

Seejärel taheti teada, kas radioloogiliste uuringute tüüpide sees esineb anomaaliaid. Selleks leiti eraldi röntgenuuringu-, ultraheliuuringu-, kompuutertomograafia uuringute- ja magnetresonantstomograafia uuringute üldist teenusekasutust kirjeldavad mudelid aastate 2010 – 2017 andmete põhjal. Tuli välja, et radioloogilised uuringud on väga varieeruvad. Kui luua lihtne üldine mudel, millega aastast või nädalast teenusekasutust võrrelda, siis ei ole see väga täpne suure varieeruvuses tõttu.

Analüüsi käigus selgus, et riigipühade ajal on teenusekasutuses märgatavad langused. Seetõttu soovib autor teada, kas oleks võimalik luua täpsem mudel, kus arvestatakse nädalas olevate tööpäevade arvu ning seeläbi vähendada varieeruvust. Näitena tuuakse kaheksas nädal, millal on Eesti Vabariigi aastapäev. Sellele päevale ei planeerita tavapäraseid uuringuid ning mudelit luues saaks arvestada, et kaheksandal nädalal on 4 tööpäeva, millest üks on lühendatud. Kuna riigipüha võib olla ka nädalavahetusel, siis riigipühale eelnevat lühendatud tööpäeva arvesse ei võeta ning kaheksandal nädalal olevate tööpäevade arvuks loetakse 4. Kui luua mudel, kus

suhtelise sageduse  $\hat{p}$  asemel on

$$\frac{\hat{p}}{\text{tööpäevade arv}}, \quad (2.2)$$

siis võiks eeldada, et eelnevalt nähtav varieeruvus väheneb. Kaheksanda nädala keskmine suhteline sagedus  $\hat{p} = 0,0179$  ja  $\frac{\hat{p}}{\text{tööpäevade arv} = 4} = 0,0045$ . Leitakse ka nädalate 2 – 7 ja 9 – 17 keskmine suhteline sagedus  $\hat{p} = 0,0208$  ja ühe tööpäeva keskmine suhteline sagedus  $\frac{\hat{p}}{5} = 0,0042$ . On näha, et varieeruvus väheneb ja tööpäevade arvu arvesse võttes mudel oleks täpsem.

Kui haiglad soovivad anomaaliaid tuvastada teenusepõhiselt, siis autori arvates oleks parim variant võrrelda teenusekasutust selle klasteri üldise teenusekasutusega, kuhu uuritav tunnus kuulub. Selleks, et mudel oleks täpsem, tuleks eelnevalt leitud klasterid radioloogia eriala arstil üle vaadata ning seletada meditsiinilisest aspektist, miks need uuringud ühes klasteris on. Enne uuringute klasterdamist tuleks leida nädalatele kaalud eelnevalt kirjeldatud meetodi abil, kus leitakse nädala keskmine sagedus võttes arvesse ka antud nädalas olnud tööpäevi. Lisaks tuleks aastate lõikes keskmist leides kasutada kärbitud keskmist ehk eemaldada minimaalne ja maksimaalne vaatlus. Seejärel klasterdaks autor uuringud ning leiaks igale klasterile keskmise suhtelise sageduse, mis vastaks üldmudelile.

# 3 Korduvate uuringute tuvastamine

## 3.1 Andmestiku ülevaade

Andmestik koosneb Eesti Haigekassa poolt 2013. aastal väljamakstud raviarvetest (enamus raviarvetest on ka koostatud aastal 2013, kuid esineb ka arveid aastast 2012). Kokku on andmestikus 348 342 rida ning 10 tunnust. Kõik andmed on pseudonümiseeritud ja andmeridu pole võimalik inimestega kokku viia. Andmestikust valitakse kõik read, mis sisaldavad radioloogilisi uuringuid ning alles jääb 154 761 rida. Tabelis 3.1 on näha korduvuuringute tuvastamiseks kasutatava andmestiku näidisrida.

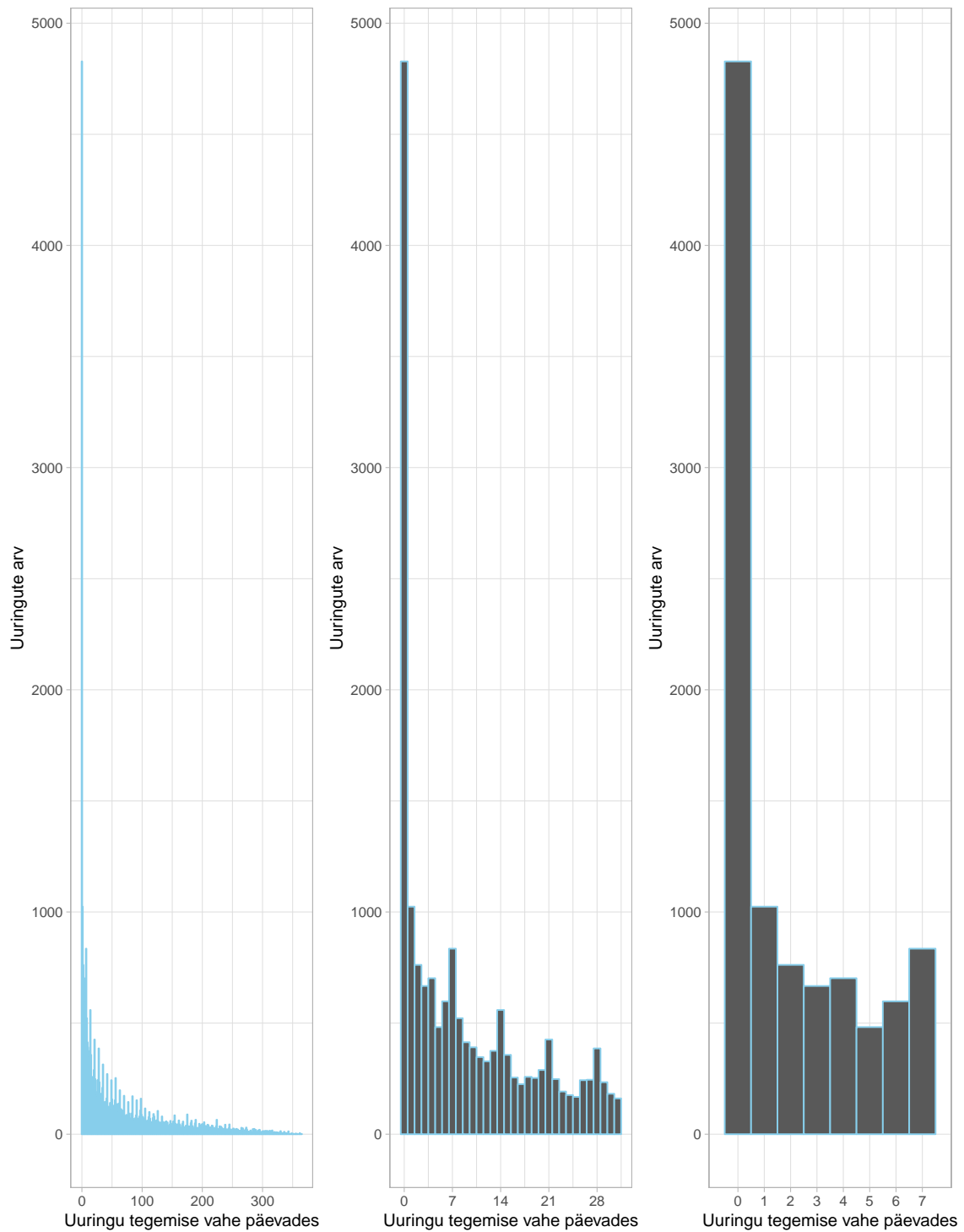
Tabel 3.1: Näidisrida radioloogiliste uuringute andmestikust koos diagnoosidega.

18	Patsiendi ID
100943	Arve ID
7903	Teenusekood
Röntgeniülesvõte rindkere piirkonnast (üks ülesvõte)	Teenuse nimi
2013-07-11	Teenuse osutamise kuupäev
P	Ravi tüüp
J20	Diagnoosikood
Äge bronhiit	Diagnoosi nimi
15	Lõpetamise staatus
Paranemine või tervistumine	Lõpetamise põhjus

## 3.2 Korduvate uuringute tuvastamine

Esmalt soovitakse leida kõik teenused, mida on tehtud patsiendile mitu korda. Jäetakse alles kõik read, kus samale patsiendile on tehtud sama uuring ning li-

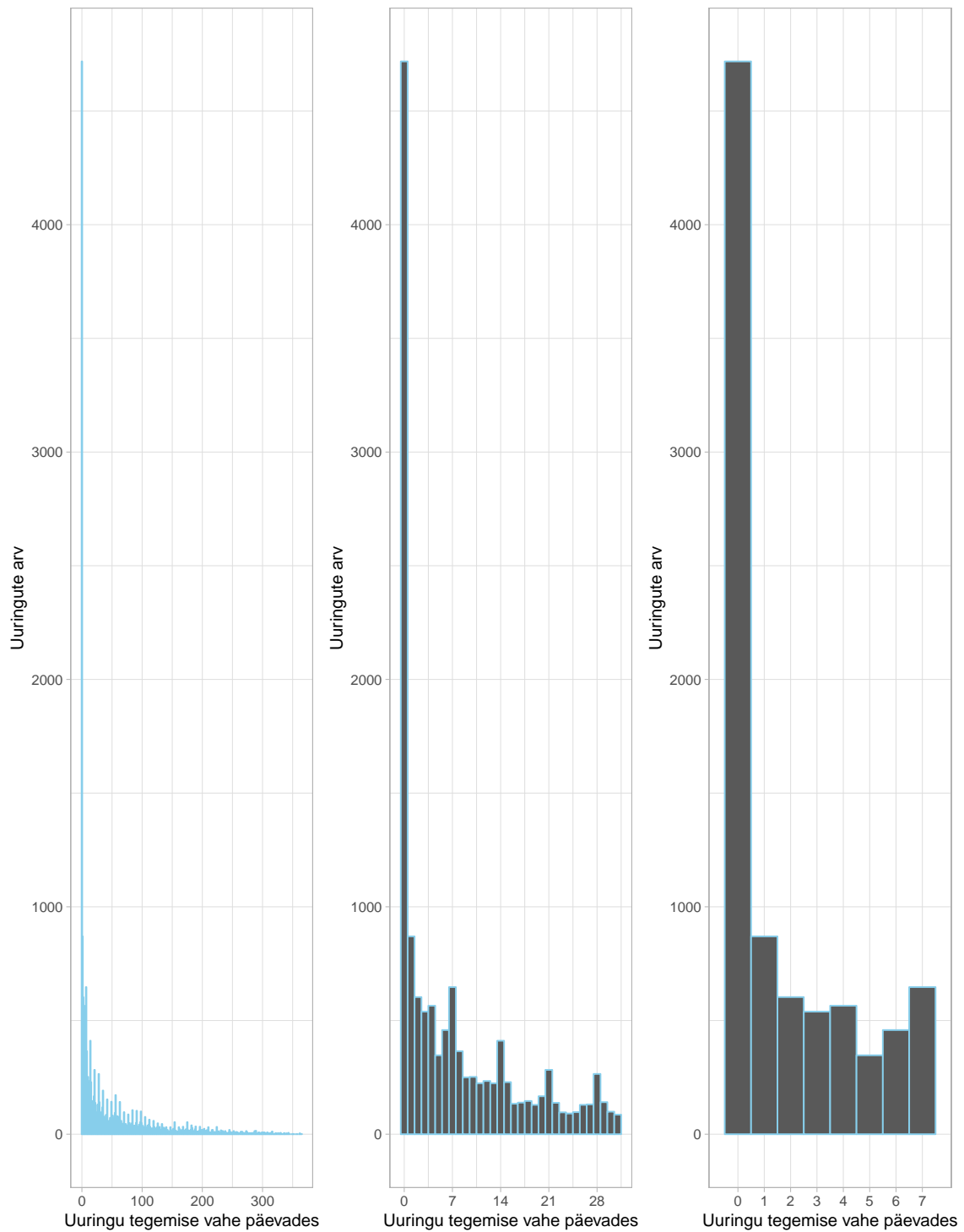
satakse ka järgneva uuringu toimumise kuupäev. Neid ridu on kokku 32 750, seega moodustavad korduvad uuringud algsest andmestikust 21,2%. Edasi valitakse need uuringud, mida on kuu aja jooksul korratud. Neid on kokku 17 132, mis moodustab 52,3% kõikidest korduvatest uuringutest ning 11% kõikidest uuringutest. Seejärel vaadatakse, kui palju korduvuuringuid tehakse ühel nädalal. Ühe nädala jooksul korratavoid uuringuid on andmestikus 9898, ehk 6,4% algsest andmestikust ja 30% kõikidest korduvuuringutest. Autor märkab, et enamuse korduvuuringutest moodustavad samal päeval tehtud uuringud. Andmestikus on samal päeval tehtud korduvaid uuringuid 4828 tükki. See moodustab kõikidest korduvuuringutest 14,7% ja 3,1% kõikidest uuringutest.



Joonis 3.1: Korduvuuringute jaotus aasta, kuu ja nädala lõikes

Eelnevalt jäeti alles kõik vaatlused, kus samale patsiendile oli tehtud sama uuring. Kuna andmestikus on ka diagnoosikood, siis võetakse see korduvate uuringute tuvastamisel arvesse. Kui patsient läheb uue arsti juurde ja arst arvab, et patsiendi diagnoos on midagi muud, kui eelnevalt määratu, siis tehakse enamasti ka uus uuring.

Korduvaid uuringuid, mis on sama diagnoosiga, on andmestikus kokku 20 870. See moodustab algsest andmestikust 13,5%. Samal kuul tehtud korduvaid uuringuid on 13 207, mis moodustab korduvatest uuringutest 63,3% ning kõikidest uuringutest 8,5%. Samal nädalal tehakse 8747 korduvat uuringut, mis moodustab korduvuuringutest 41,9% ja kõikidest uuringutest 5,7%. Jällegi on näha, et suurima osa korduvatest uuringutest moodustavad samal päeval tehtud korduvuuringud, mida on kokku 4718. See on 22,6% kõikidest korduvuuringutest ning 3% kõikidest uuringutest.



Joonis 3.2: Sama diagnoosiga korduvuuringute jaotus aasta, kuu ja nädala lõikes

Eelnevalt ei olnud määratud, mis tüüpi arsti juures patsient käis. Nüüd keskendutakse ainult nendele uuringutele, mida korratakse teist tüüpi arsti juures.

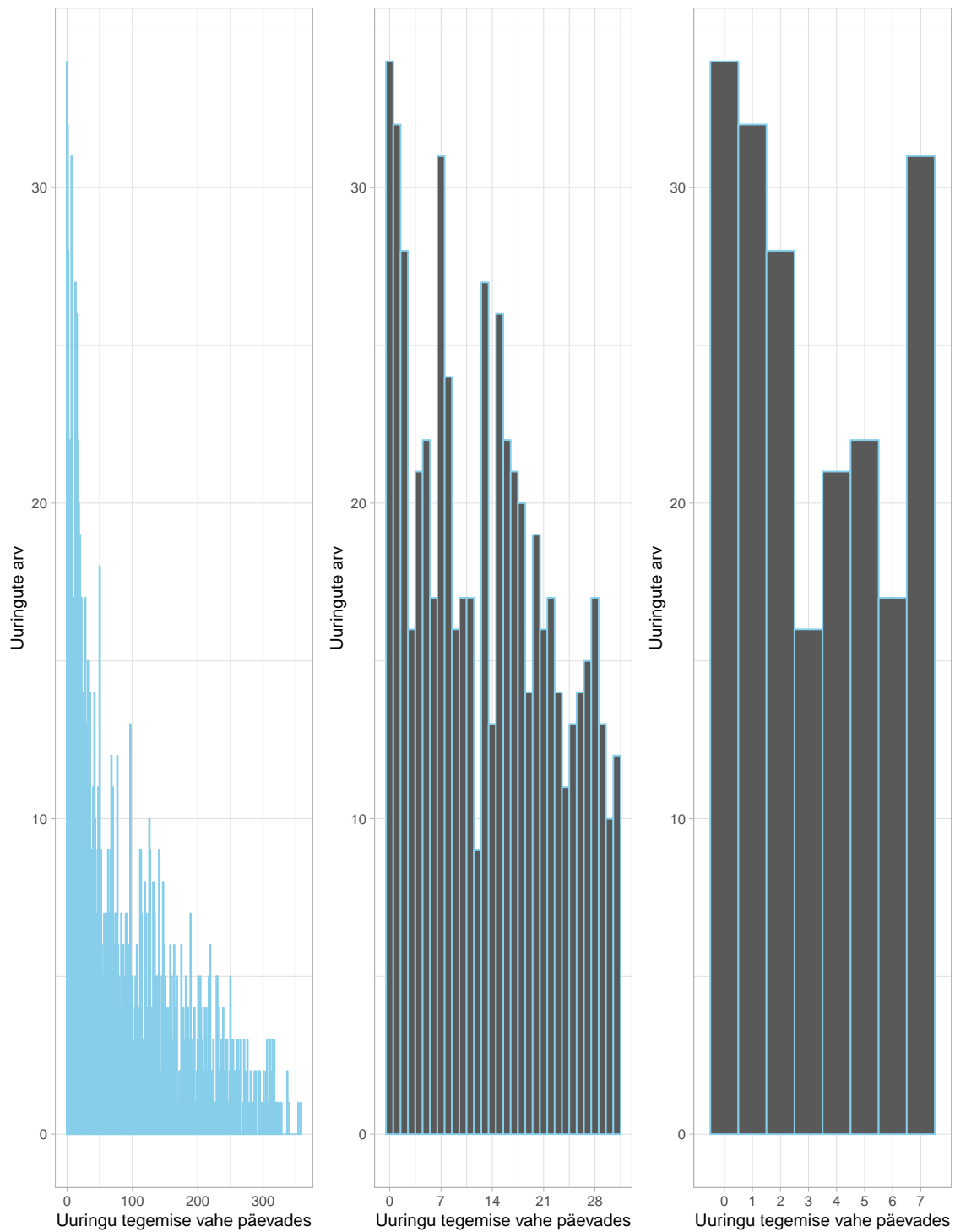
Võimalikud liikumised on:

- perearsti juurest ambulatoorsesse või statsionaarsesse vastuvõttu,

- ambulatoorsest või statsionaarsest edasi perearsti vastuvõtule.

Edaspidi nimetatakse ambulatoorset ja statsionaarset vastuvõttu eriarsti vastuvõtuks. Perearsti ja eriarsti juures antavad diagnoosid ei ole alati samad. Seetõttu korduvuuringute tuvastamisel diagnoosiga ei arvestata. Eelnevalt leiti, et korduvaid uuringuid oli kokku 32 750.

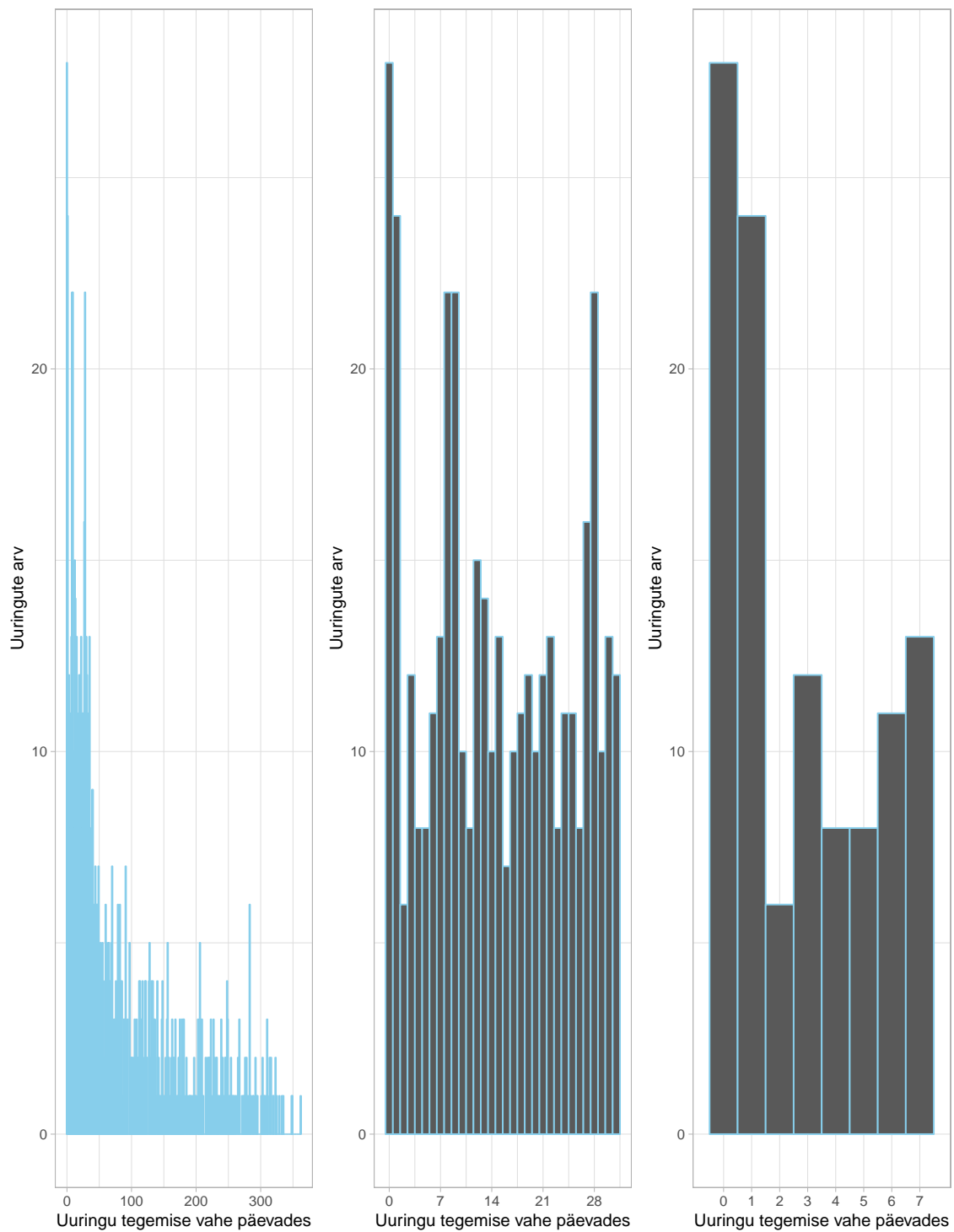
Esimesena vaadeldakse juhtu, kus esmane uuring tehakse perearsti juures ning järgnev eriarsti juures. Selliseid uuringuid on kokku 1653, mis on 5% kõikidest korduvuuringutest. Nendest uuringutest 598 on tehtud väiksema vahega kui üks kuu. Kõikidest leitud uuringutest, kus esmane uuring tehti perearsti juures ja järgmine eriarsti juures, moodustab see 36,2%. Samal nädalal korratud uurinuid on 201 tükki, mis on 12,2% leitud korduvuuringutest. Kui eelnevalt olid kõige sagedasemad samal päeval korratud uuringud, siis erinevate arstide juures käies sageduse erinevus nii suur ei ole. Samal päeval korratud uuringuid on 34, ehk 2% kõikidest korduvatest uuringutest, mis on tehtud erinevate arstide juures.



Joonis 3.3: Perearsti juures tehtud uuringule järgnenud eriarsti vastuvõtul tehtud korduvuuringute jaotus aasta, kuu ja nädala lõikes

Nüüd vaadeldakse uuringuid, mis esmalt sooritati eriarsti juures ning seejärel tehti sama uuring perearsti juures. Neid korduvaid uuringuid on 1014 tükki, ehk 3,2%. Samal kuul tehtud uuringuid on 410, mis moodustavad 40,4% korduvatest uuringu-

test. Nädala jooksul korratud uuringuid on 110, ehk 10,8% ja samal päeval tehtud uuringuid on 2,8% ehk 28 tükki.



Joonis 3.4: Eriarsti juures tehtud uuringule järgnenud perearsti vastuvõtul tehtud korduvuuringute jaotus aasta, kuu ja nädala lõikes

Tabel 3.2: Korduvuuringute arv aastas, kuus ja nädalas.

	Kõik korduvad uuringud	Sama diagnoosiga korduvad uuringud	Eriarst kordab perearsti uuringut	Perearst kordab eriarsti uuringut
Korduvuuringute arv aastas	32 750	20 870	1653	1014
Osakaal kõikidest korduvatest uuringutest	100%	63,7%	5%	3,1%
Osakaal kõikidest uuringutest	21,2%	13,5%	1,1%	0,7%
Korduvuuringute arv kuus	17 132	13 207	598	410
Osakaal kõikidest korduvatest uuringutest	52,3%	40,3%	1,8%	1,3%
Osakaal kõikidest uuringutest	11,1%	8,5%	0,4%	0,3%
Korduvuuringute arv nädalas	9898	8747	201	110
Osakaal kõikidest korduvatest uuringutest	30,2%	26,7%	0,6%	0,3%
Osakaal kõikidest uuringutest	6,4%	5,7%	0,13%	0,07%
Korduvuuringute arv päevas	4828	4718	34	28
Osakaal kõikidest korduvatest uuringutest	14,7%	14,4%	0,1%	0,09%
Osakaal kõikidest uuringutest	3,1%	3%	0,02%	0,02%

Autor uurib edasi neid korduvuuringuid, mida on korratud ühe nädala jooksul. Neid on kokku 9898. Korduvuuringuid, kus eriarst kordab perearsti tehtud uuringut, on 201. Korduvuuringuid, kus perearst kordab eriarsti tehtud uuringut, on 110. Juhumeid, kus patsient suunatakse teise haiglasse ning talle seal uuesti sama uuring tehakse, on 492. Need korduvuuringud on seletatavad, sest teine arst võib kahtlustada muud diagnoosi ning seetõttu uue radioloogilise uuringu teha. Lisaks on nende osakaal samal nädalal korratud uuringutest üsna väike - 8,2%. Enamuse korduvuuringutest moodustavad juhtumid, kus patsiendile tehakse ühe raviarve sees korduvuuring - neid on 7164. Kuna sinna sisse kuuluvad ka uuringud, mida on vaja mitu tükki teha, näiteks mammograafia ja silmapõhja optiline koherentne tomograafia, siis need loetakse põhjendatud korduvateks uuringuteks. Jättes välja eelnimetatud radioloogilised uuringud, mida tehakse korraga kaks, siis jääb ühe raviarve sees tehtud korduvaid uuringuid alles 6238. Selleks, et mõista, miks antud uuringud on korduvad, tuleks kaasata meditsiiniline ekspert. Leitud korduvate uuringute arv on kõikidest uuringutest 4%, mis ei ole nii suur osakaal, et korduvad uuringud oleks suur probleem.

# Kokkuvõte

Töö eesmärgiks oli tuvastada radioloogiliste uuringute tegemises anomaaliaid. Sooviti näha, kas uuringute tegemisel on ajalisi anomaaliaid ning tuvastada korduvaid uuringuid.

Töö esimeses osas loodi ajaliste anomaaliade tuvastamiseks üldist teenusekasutust kirjeldav mudel. Radioloogiliste uuringute tegemist võrreldi loodud mudeliga ja leiti, et üldiselt tehakse radioloogilisi uuringuid mudelile vastavalt. Seejärel uuriti eraldi röntgen-, ultraheli-, kompuutertomograafia- ja magnetresonantstomograafia uuringute teenusekasutust ja loodi üldist teenusekasutust kirjeldav mudel. Võrreldes teenusekasutust üldmudelitega leiti, et radioloogilistel uuringutel leiduvad ajutised anomaaliaid.

Töö teise osa eesmärk oli tuvastada korduvaid radioloogilisi uuringuid. Selleks leiti kõik tehtud korduvad uuringud, seejärel leiti tehtud korduvad uuringud, mis olid sama diagnoosiga ning nende osakaalud kõikidest tehtud uuringutest ning kõikidest korduvatest uuringutest. Lisaks leitakse eri tüüpi arstide korratud uuringute arv ja osakaal. Töös selgub, et enamus korduvatest uuringutest on sama diagnoosiga.

# Kasutatud kirjandus

Cort J. Willmott, Kenji Matsuura (2005). “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance”. *Climate Research* 30.1, l. 79–82.

Gareth James Daniela Witten, Trevor Hastie Robert Tibshirani (2013). *An Introduction to Statistical Learning with Applications in R*. Springer, l. 390–396.

James Nol, Godfrey Isouard Jerzy Mirecki (2005). “Uncovering the causes of unnecessary repeated medical imaging examinations, or part of, in two hospital departments”. *The Radiographer* 52.3, l. 26–31.

Siegel, Sidney (1956). *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, l. 36–37.

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

Mina, Tuule Tamme,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Radioloogiliste korduvuuringute tuvastamine globaalsel skaalal Eesti Haigekassa näitel“, mille juhendaja on Sven Laur, 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.

*Tuule Tamme*  
**08.05.2019**