

TARTU ÜLIKOOL  
Arvutiteaduse instituut  
Andmeteaduse õppekava

**Paula Marie Loopere**  
**Laboris mõõdetavate analüütide**  
**referentsväärtuste määramine Eesti**  
**rahvastiku näitel**

**Magistritöö (15 EAP)**

Juhendajad:  
Jaak Vilo, PhD  
Sulev Reisberg, PhD

Tartu 2025

# **Laboris mõõdetavate analüütide referentsväärtuste määramine Eesti rahvastiku näitel**

## **Lühikokkuvõte:**

Laboratoorsed analüüsid on olulised patsiendi tervisliku seisundi hindamisel ning neil on suur osa raviotsuste tegemisel. Analüüside tulemuste tõlgendamiseks on vajalikud referentsväärtused, mis arvestaksid ka kohaliku populatsiooni eripärasid. Eestis kasutatakse sageli tootjate poolt määratud referentsväärtuseid. Need ei pruugi olla kohased, sest puudub info nende loomisel kasutatud populatsiooni kohta. Magistritöö eesmärk on määrata Eesti populatsiooni põhjal referentsväärtused kolmele kliinilise keemia ja immunoloogia analüüdile: naatrium, kreatiniin ja ferritiin. Töö tulemusena määrati naatriumile üks üldine referentsvahemik. Kreatiniinile leiti soospetsiifiline ning ferritiinile soo- ja vanusespetsiifilised referentsväärtused. Töös määratud referentsväärtused on üldiselt kooskõlas hetkel kasutusel olevate väärtustega, kuid ilmnas erinevusi. Lisaks testiti väljatöötatud meetodit uuemate andmete peal ning magistritöö käigus valmis kasutajaliides, mis võimaldab nii andmete visualiseerimist kui ka referentsväärtuste määramist.

**Võtmesõnad:** Referentsväärtused, laborianalüüsid, naatrium, kreatiniin, ferritiin

**CERCS:** B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika; P160 Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika

# **Calculation of reference values for laboratory measurements based on the Estonian population**

## **Abstract:**

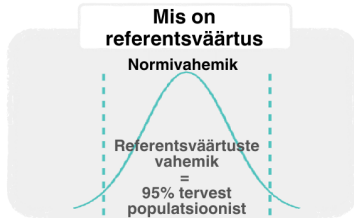
Laboratory tests are important when evaluating a patient's health and play a major role in making treatment decisions. Reference values are necessary to interpret the results of analyses, which also take into account the special features of the local population. In Estonia, often are used the reference values which are determined by manufacturers. These may not be the most appropriate, because there is no information about the population used to create them. The aim of this master's thesis is to calculate reference values for three clinical chemistry and immunology analytes based on the Estonian population. These analytes are sodium, creatinine and ferritin. As a result of the work, one general reference range was determined for sodium. Gender-specific reference values were found for creatinine and gender- and age-specific reference values for ferritin. The reference values determined in the work are generally consistent with the values currently in use, but some differences were found. The developed method was tested on newer data. In addition, a user interface was developed during the master's thesis, which allows both data visualization and determination of reference values.

**Keywords:** Reference values, laboratory analysis, sodium, creatinine, ferritin

**CERCS:** B110 Bioinformatics, medical informatics, biomathematics, biometrics; P160 Statistics, operation research, programming, actuarial mathematics

# Visuaalne kokkuvõte

## Laboris mõõdetavate analüütide referentsväärtuste määramine Eesti rahvastiku näitel



**Probleem**

Tootjate referentsväärtused

Eesti populatsiooni väärtused

**Terviseandmed**

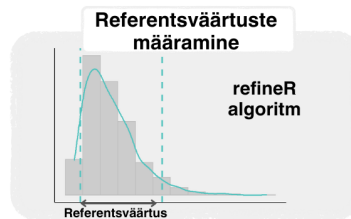
**RITA MAITT andmestik**  
Eesti 10% juhuvalim tervise infosüsteemist aastatel 2012 kuni 2019

**Uuritavad analüüdid**

Naatrium

Kreatiniin

Ferritiin



**Tulemused**

**Leitud referentsväärtused**  
Naatrium: üldine  
Kreatiniin: soospetsiifiline  
Ferritiin: soo- ja vanusespetsiifiline

Metoodika laiendatav teistele laborianalüüsidele

**Kasutajaliides**

Autor: Paula Marie Loopere  
Juhendajad: Jaak Vilo (PhD), Sulev Reisberg (PhD)  
Andmeteadeus (MSc), 2025

#UniTartuCS



TARTU ÜLIKOOL  
arvutiteaduse instituut

# Visual abstract

**Calculation of reference values for laboratory measurements  
based on the Estonian population**

<p><b>What is reference value?</b></p> <p>Normal range Reference value range = 95% of the healthy population</p>	<p><b>Problem</b></p> <p>Manufacturer's reference values <math>\neq</math> Estonian population values</p>	<p><b>Data</b></p> <p><b>RITA MAITT dataset</b> 10% random sample from the national health databases 2012-2019 in Estonia</p>
<p><b>Analytes</b></p> <p>Sodium Creatinine Ferritin</p>	<p><b>Calculation of reference values</b></p> <p>refineR algorithm</p> <p>Reference value</p>	<p><b>Results</b></p> <p><b>Determined reference values</b> Naatrium: general Kreatiniin: gender-specific Ferritiin: gender- and age-specific Methodology applicable to other lab tests</p> <p><b>User interface</b> </p>

Author: Paula Marie Loopere  
Supervisors: Jaak Vilo (PhD), Sulev Reisberg (PhD)  
Data Science (MSc), 2025

#UniTartuCS

UNIVERSITY OF TARTU  
Institute of Computer Science

# Sisukord

1. Sissejuhatus .....	8
2. Kirjanduse ülevaade .....	10
2.1 Vajadus ühiste referentsväärtuste järele .....	10
2.2 Referentsväärtuste määramise meetodid .....	10
2.3 Uuritavad analüüdid .....	11
3. Andmed ja metoodika .....	14
3.1 Andmed .....	14
3.2 Metoodika .....	15
3.2.1 Wilcoxon'i astaksummatest .....	15
3.2.2 Boxi-Coxi teisendus .....	15
3.2.3 RefineR algoritm .....	16
3.2.4 Eetikakomitee luba .....	18
4. Tulemused .....	19
4.1 Naatrium .....	19
4.2 Kreatiniin .....	21
4.3 Ferritiin .....	24
4.4 Lisa-analüüs tervise infosüsteemi andmetel .....	27
4.5 Kasutajaliides .....	30
5. Arutelu .....	31
5.1 Naatrium .....	31
5.2 Kreatiniin .....	32
5.3 Ferritiin .....	32
5.4 Kasutajaliides .....	33
5.5 Töö tugevused, nõrkused ja edasised sammud .....	34
6. Kokkuvõte .....	36
7. Tänuõnad .....	38
Viited .....	39
Lisad .....	41
I. Referentsväärtused .....	41
II. Kasutajaliidese avalehe vaheleht .....	43
III. Kasutajaliidese andmete ülevaate vaheleht .....	44
IV. Kasutajaliidese vanuse ja soo analüüsi vaheleht .....	45

V. Kasutajaliidese referentsväärtuste määramise vaheleht .....	46
VI. GitLabi projekt .....	47
Litsents .....	48

# 1. Sissejuhatus

Patsiendi tervise uurimisel tehakse tema seisundi hindamiseks erinevaid laboratoorseid analüüse. Iga analüüsi käigus määratakse kvantitatiivselt ühe või mitme analüüdi kogus. Nendel analüüsidel on suur roll: umbes 70% raviotsustest toetub laborianalüüsides tulemustele (Forsman, 1996). Bioloogiliselt võivad mõõdetavad analüüdid aga sõltuda nii soost kui ka vanusest ja lisaks erineda rassiti või populatsiooniti. Seetõttu on tulemuste tõlgendamiseks vaja võrdlemiseks kõrvale ka vastava analüüsi referentsväärtust (Ceriotti ja Henny, 2008). Näiteks Eestis on praegu täiskasvanutel mõõdetava naatriumi referentsväärtus 136–145 mmol/L, mis tähendab, et sellest vahemikust madalamad või kõrgemad analüüsi tulemused võivad viidata terviserikkele (Kallion, 2011). Usaldusväärsed referentsväärtused on olulised, sest need aitavad tervisehoiutöötajatel otsustada, kas analüüsi tulemus jääb normi piiridesse või viitab võimalikele terviseprobleemidele.

Eesti laborites kasutatakse referentsväärtustena labori mõõtmeseadmete tootjate poolt määratud referentsväärtuseid. Kahjuks ei ole tootjate referentsväärtuste puhul teada nende loomisel uuritud populatsioonid, samuti võivad puududa usaldusväärsed väärtused ka laste ja vanemate inimeste kohta (Bohn jt, 2023). Lisaks erinevate mõõtmeseadmete ja kemikaalide kasutamisele võivad analüüsides mõõtmistulemused erineda isegi siis, kui kasutatakse sama analüüsimeetodit. See võib põhjustada referentsväärtuste varieerumist laborite vahel, mis raskendab tulemuste võrreldavust ja nende järjepidevust (Bohn jt, 2023). Alternatiivse lahendusena võib kaaluda suurandmete baasil välja arvatud rahvastikupõhiseid referentsväärtusi. Need põhinevad laborite igapäevatoos kogutud mõõtmistulemustel, mitte piiratud valimiga kliinilistel uuringutel, mis on läbi viidud seoses seadme turule toomisega. Näiteks on seda edukalt rakendatud Kanadas, kus professor Khosrow Adeli juhtimisel määrati kohaliku populatsiooni põhjal referentsväärtused 16 analüüdile (Bohn jt, 2023). Sarnase lähenemise rakendamine Eesti populatsioonil aitaks parandada referentsväärtuste täpsust kohaliku populatsiooni eripärasid silmas pidades.

Magistritöö eesmärk on määrata suurandmetel põhinevad referentsväärtused Eesti rahvastiku jaoks. Referentsväärtused töötatakse välja kaudsel meetodil olemasolevate terviseandmete põhjal, kasutades tervise infosüsteemi andmekogu. Eeldatavasti esineb mõnede analüütide puhul nii vanuselisi kui ka soost tingitud erinevusi, seetõttu leitakse soo- ja vanusespetsiifilised väärtused. Samuti ei pruugi kõikide analüütide tulemuste ühtlustamine olla võimalik, kuna erinevate tootjate analüüsimeetodid võivad liialt erineda. Töös keskendutakse kliinilise keemia ja immunoloogia automaatanalüsaatoritel mõõdetavatele analüütidele. Metoodika väljatöötamiseks on välja valitud kolm analüüti: naatrium (seerumis/plasmas), kreatiniin (seerumis/plasmas)

ja ferritiin (seerumis/plasmas). Referentsväärtuste arvutamiseks ning visualiseerimiseks on magistritöö raames loodud interaktiivne kasutajaliides, kuhu saab ise ka vajalikke andmeid sisestada ja uusi referentsväärtusi välja arvutada. Lisaks rakendatakse välja töötatud metoodikat ka uuemates terviseandmetes, et hinnata selle stabiilsust ja usaldusväärsust.

Magistritöö on jagatud kuueks peatükiks. Teises peatükis antakse ülevaade referentsväärtustest ning nende arvutamise meetoditest, samuti töös uuritavatest analüütidest ning praegu Tartu Ülikooli Kliiniku laboris, SYNLABis ja Kanada laboris kasutatavatest referentsväärtustest. Töö kolmandas osas kirjeldatakse referentsvahemike arvutamiseks kasutatavaid statistilisi meetodeid ja algoritmi. Neljandas peatükis antakse ülevaade töö tulemustest. Analüüsitakse andmete jaotust soo ja vanuse lõikes ning määratakse kolmele analüüdile referentsväärtused Eesti rahvastiku näitel. Samas peatükis kirjeldatakse ka metoodika testimist uuematel andmetel. Viiendas peatükis on tulemuste arutelu ja antakse ülevaade loodud kasutajaliidesele. Töö lõpus on esitatud magistritöö kokkuvõtte, järeldused ja välja toodud ka edasised uurimissuunad.

Töö raames soovitakse kontrollida järgmisi hüpoteese:

1. Eesti populatsioonil arvutatud referentsväärtused erinevad tootjate poolt määratud referentsväärtustest.

*Populatsioonipõhised referentsväärtused võivad erineda tootjate esitatud väärtustega, sest nendes ei ole arvestatud rahvastiku eripäradega.*

2. Sugu ja vanus mõjutavad referentsväärtuste määramist.

*Analüütide jaotustes võib esineda märkimisväärsed erinevusi nii soo kui ka vanuse lõikes. Sellisel juhul on vaja määrata eraldi referentsväärtused erinevatele demograafilistele rühmadele.*

## **2. Kirjanduse ülevaade**

Käesolevas peatükis kirjeldatakse, kuidas on referentsväärtuseid maailmas varem määratud. Antakse ülevaade erinevatest analüüsimeetoditest, mida kasutatakse referentsväärtuste määramiseks. Seejärel antakse ülevaade antud töös uuritavatest analüütidest.

### **2.1 Vajadus ühiste referentsväärtuste järele**

Referentsväärtuste määramine on olnud oluline erinevates teadus- ja tehnoloogiavaldkondades, alustades sotsiaalteadustest kuni loodusteadusteni. Mitmetes valdkondades on olemas standardid ja normid, mille suhtes saab võrrelda uusi andmeid. Erinevatel teadusaladel kasutatakse selleks oma spetsiifilisi termineid, lähenemisviise ja määramismeetodeid.

Terviseandmete puhul hakati referentsväärtusi teaduslikult defineerima 20. sajandi teisel poolel (Geffré jt, 2009). Enne seda viidati nendele kui normaalsetele analüüsi tulemustele. Tänapäeval viitab referentsväärtus vahemikule, millesse jääb 95% tervete inimeste mõõdetud väärtustest (Geffré jt, 2009). Meditsiinis aitavad referentsväärtused hinnata, kas inimese tervisenäitajad on korras või viitavad võimalikele terviseprobleemidele. Täpselt määratud referentsväärtused on seega tähtsad haiguste ennetamisel, diagnoosimisel ning ravi jälgimisel (Geffré jt, 2009).

### **2.2 Referentsväärtuste määramise meetodid**

Populatsioonipõhiste referentsväärtuste määramiseks on olemas kaks peamist meetodit: otsene ja kaudne meetod.

Otsene meetod järgib kliiniliste ja laboratoorsete standardite instituudi EP28-A3c juhiseid (Clinical and Laboratory Standards Institute, 2010). Otsese meetodi puhul analüüsitakse ainult tervete inimeste valimit. Tavaliselt valitakse osalejad kindlate kriteeriumite alusel, näiteks soo ja vanuse järgi. Neist võetakse juhuvalim, mis esindab tervete inimeste populatsiooni. Seejärel tehakse statistiline analüüs, et määrata keskmine 95% referentsvahemik. Teisisõnu leitakse 2,5% ja 97,5% protsentiilid, mis on vastavalt alumine ja ülemine referentsväärtus.

Kaudne meetod põhineb labori olemasolevatel kliinilistel andmetel, mis sisaldavad sageli ka juba haigete inimeste tulemusi. Seetõttu tuleb enne analüüsi andmeid eeltöödelda, see tähendab eemaldada kordused ja piirata korduvaid mõõtmisi. Esiteks hõlmab andmete filtreerimine näiteks erakorralise meditsiini patsientide tulemuste eemaldamist. Kui patsiendil on tehtud mitu sama analüüsi teatud aja jooksul, näiteks aastas, siis neid mõõtmisi andmetes ei kaasata. Kuna analüüside tulemused võivad olla tugeva kaldega ja ei järgi normaaljaotust, siis tihti

on vajalik andmete teisendamine. Seetõttu kasutatakse sageli mitteparameetrilisi meetodeid referentsväärtuste arvutamiseks (Haeckel jt, 2023).

Mõlemal meetodil on oma eelised ja puudused. Otsene meetod võib olla täpsem, kuid andmete kogumine on kulukam ja aeganõudvam. Kaudne meetod on kiirem, kuid vajab keerulisemat statistilist analüüsi ning tulemused sõltuvad andmete kvaliteedist (Jones jt, 2019). Meetodi valik sõltub kasutatavatest andmetest ning labori või uuringu eesmärkidest. Tavaliselt vastutab labor ise referentsvahemike valideerimise või määramise eest, võttes arvesse piirkonna populatsiooni eripärasid ja kasutatavaid analüüsimeetodeid.

Kaudseid meetodeid on mitmeid, millest levinumad on Hoffmani, refineR ja KOSMIC meetodid. Hoffman-i meetod on üks vanimaid kaudseid meetodeid, mis on hästi lihtne ja kiire (Meyer jt, 2023). Kahjuks eeldab see meetod andmete normaaljaotust, mida tihti mõõtmistulemustes ei esine. KOSMIC ja refineR algoritme peetakse paremaks valikuks, sest need ei eelda normaaljaotust ja suudavad ise teisendada andmed normaaljaotuse lähedaseks jaotuseks (Ammer jt, 2021). KOSMIC algoritmi jõudlus halveneb märkimisväärselt, kui andmestikus esineb rohkelt haigustest mõjutatud tulemusi (Ammer jt, 2021). RefineR meetod on üks uusimaid algoritme.

Käesolevas magistritöös kasutatakse refineR algoritmi, kuna see sobib hästi kaudseks referentsväärtuste määramiseks ega eelda andmete normaaljaotust. Algoritm suudab automaatselt teisendada andmeid normaaljaotuse kuju sarnaseks. See tagab täpsemad ja usaldusväärsemad tulemused isegi keerulistes andmekogumites, nagu seda terviseandmed tihtipeale on. Esmalt tuvastab algoritm andmetes peamise tipu, mis kõige tõenäolisemalt esindab tervete isikute analüüsitulemusi. Seejärel kasutab algoritm iteratiivset lähenemist, et leida kõige paremini sobiv jaotuse mudel, mis aitab eristada haigusest mõjutatud tulemusi. Optimaalne mudel valitakse selle järgi, kui hästi see sobitub algandmetega. Referentsväärtused määratakse hinnatud tervete isikute tulemusi kirjeldava jaotuse protsentiilide põhjal. Algoritmi täpsem metoodika on kirjeldatud peatükis 3.2. See algoritm avaldati 2021. aastal teadusartiklis "refineR: A Novel Algorithm for Reference Interval Estimation from Real-World Data", mille autorid on Tatjana Ammer, André Schützenmeister, Hans-Ulrich Prokosch, Manfred Rauh, Christopher M. Rank ja Jakob Zierk. (Ammer jt, 2021)

### **2.3 Uuritavad analüüdid**

Referentsväärtused leitakse kliinilise keemia ja immuunanalüüsi automatiseeritud meetoditel määratavatele analüütidele. Nendeks analüütideks on naatrium, kreatiniin ja ferritiin.

Naatrium on vajalik mineraal organismi vedelike tasakaalu hoidmiseks, mida leidub peamiselt rakuvälises vedelikus (Kallion, 2011). Naatrium jõuab kehasse toiduga ja väljub peamiselt neerude kaudu, aga ka higistamisega ning selle taset reguleerivad hormoonid (Kallion, 2011). Naatrium aitab hoida vedeliku tasakaalu, happe-aluse tasakaalu ja närvide ning lihaste tööd (Kallion, 2011). Vere naatriumisaldus näitab pigem naatriumi ja vedeliku vahelist tasakaalu kui kogu naatriumikogust kehas (Kallion, 2011). Kuna naatriumi bioloogiline ja analüütiline varieeruvus on väga väike, siis peaksid töös leitavad referentsväärtused peaaegu täielikult kokku langema hetkel laborite poolt kasutatavatega. Tabelis 1 on esitatud uuritavate analüüside referentsväärtused, täpsemalt Kanadas ja kahes Eesti laboris kasutusel olevad väärtused. Eestis on täiskasvanutel kasutusel naatriumi referentsväärtus 136–145 mmol/L. Mujal võivad väärtused olla erinevad, näiteks Kanadas on hetkel kasutusel vahemik 136–146 mmol/L, mis on üsna sarnane, kuid ülemise piiri võrra veidi suurem Eestis kasutusel oleva referentsväärtusega.

Tabel 1. Analüüside referentsväärtused Kanada meditsiininõukogust (Medical Council of Canada, 2023) ning kahest Eesti laborist: Tartu Ülikooli Kliinikumi (TÜK) Ühendlabori (Tartu Ülikooli Kliinikumi Ühendlabor, 2025) ja SYNLAB (SYNLAB Eesti, 2024).

	<b>Naatrium</b> ( <i>mmol/L</i> )	<b>Kreatiniin</b> ( <i>μmol/L</i> )	<b>Ferritiin</b> ( <i>μg/L</i> )
<b>Kanada labor</b>	≥ 18 a 136–146	≥ 18 a 49–93 (M) ≥ 18 a 22–75 (N)	≥ 18 a 24–336 (M) ≥ 18 a 11–307 (N)
<b>TÜK Ühendlabor</b>	≥ 18 a 136–145	≥ 15 a 59–104 (M) ≥ 15 a 45–84 (N)	18–61 a 30–400 (M) 18–61 a 13–150 (N)
<b>SYNLAB</b>	≥ 19 a 136–145	≥ 19 a 64–111 (M) ≥ 19 a 50–98 (N)	≥ 19 a 28–370 (M) ≥ 19 a 10–204 (N)

Kreatiniin on lihaste töö käigus tekkiv jääkaine, mis eritub neerude kaudu ega imendu tagasi, seetõttu kasutatakse seda neerude töö hindamiseks (Reimand, 2024). Kreatiniini tase plasmas ei ole kõigil inimestel ühesugune. Kreatiniini tase võib sõltuda mitmetest teguritest, nagu vanus ja sugu, mistõttu on oluline analüüsida seda rahvastikupõhiselt (Reimand, 2024). Tabelist 1 paistab, et Eestis ja Kanadas kasutatavad referentsväärtused on veidi erinevad. Täpsemalt on Tartu Ülikooli Kliinikumi ühendlabori ja SYNLABi referentsväärtused kõrgemad kui Kanada omad. Liiga madal ja liiga kõrge kreatiniini tase võib aga viidata terviseprobleemidele.

Ferritiin on rauda sisaldav rakusisene valk, mida leidub kõige rohkem maksas, põrnas ja luuüdis ning see näitab, kui palju rauavarusid on kehas (Vaagen, 2023). Seetõttu on ferritiini taseme määramine plasmas oluline. Madal ferritiini tase viitab rauapuudusele ning liiga kõrge tase viitab põletikule (Vaagen, 2023). Uuringud on näidanud, et ferritiini referentsväärtuseid määratakse erinevalt, mis toob kaasa suure referentsväärtuste varieeruvuse laborite vahel (Kurstjens jt, 2025). See põhjustab ebakindlust tulemuste usaldusväärsuses, mistõttu oleks oluline ferritiini uurida ka rahvastikupõhiselt (Kurstjens jt, 2025). Tabelist 1 on näha, et ferritiini alumised referentsväärtused on sarnased, kuid ülemised väärtused on erinevad. Eestis on meeste ülemised piirid veidi suuremad kui Kanadas, samas kui naiste puhul on Eestis referentsväärtused märgatavalt madalamad kui Kanadas.

### **3. Andmed ja metoodika**

See peatükk annab ülevaate analüüsis kasutatavatest andmetest ja metoodikast. Selgitatakse andmete päritolu ning nende tunnuseid.

#### **3.1 Andmed**

Töö teostatakse RITA MAITT uuringu andmestikul ja tervise infosüsteemi andmestikul. Uuringu pikem eesmärk on arendada ja testida metoodikat laborianalüüside võrdlemiseks eri andmestikel, selleks on ka vastavad load olemas. Hetkel keskendutakse esimeses etapis kättesaadavatele andmetele RITA MAITT andmestikust, mis sisaldab Eesti 10% juhuvalimit tervise infosüsteemi saatekirja vastustest ja epikriisidest aastatel 2012 kuni 2019. Andmestikust võetakse isikud, kellel esineb vaadeldava analüüdi mõõtmise. Andmestikust on eemaldatud alla 18-aastaste isikute andmed. Lisaks välistati analüüsitulemused, mille puhul oli isikul samal aastal mitu sama analüüsi tulemust, kuna see võib viidata probleemsele tervises seisundile ja ebanormaalsele analüüdi väärtusele. Analüüse teostatakse analüütide kaupa.

Iga analüüdi kohta vajalikud andmed on: analüüdi nimetus, analüüsi vastuse aeg, patsiendi pseudonüüm, patsiendi sugu, patsiendi vanus, analüüsi tulemus, ühik.

Peale andmete puhastamist on naatriumi analüüside kohta 46 322 rida, millest meeste analüüse on 18 967 ja naiste analüüse 27 355. Kreatiini analüüsitulemuste kohta on 72 784 rida, millest meeste analüüsitulemusi on 28 683 ja naiste analüüsitulemusi 44 101. Viimaseks, ferritiini analüüside tulemusi on 18 656, millest meeste analüüse on 5202 ja naiste omasid üle kahe korra rohkem – 13 454.

Lisaks RITA MAITT andmestikule on kavas võrrelda ja testida kasutatud metoodikat edaspidi ka teistel andmetel. Antud töös tehakse lisa-analüüs tervise infosüsteemi andmestikul, mis sisaldab kõiki Eesti laborianalüüside tulemusi aastatel 2023 kuni 2024. Selle andmestiku eeliseks on suur andmemaht ning ajakohasus. Peale andmestiku puhastamist on naatriumi analüüsitulemusi 319 433, neist 133 020 on meeste tulemused ja 186 413 naiste tulemused. Kreatiini analüüside tulemusi on 742 887, millest meeste tulemusi on 315 891 ja naiste tulemusi 426 996. Ferritiini analüüside kohta on 325 574 rida, millest meeste analüüse on 102 194 ja naiste analüüse 223 380. Eri andmestikel leitud referentsväärtused aitavad hinnata, mil määral sõltuvad või ei sõltu referentsväärtused alusandmestike erinevustest ning kui stabiilseid tulemusi annab valitud metoodika.

## 3.2 Metoodika

Metoodika alapeatükk on jagatud kolmeks. Esimesena tutvustatakse Wilcoxon'i astaksummatesti, mida kasutatakse kahe valimi võrdlemiseks. Teiseks tutvustatakse Boxi-Coxi teisendust, mida kasutatakse andmete normaliseerimiseks. Viimases alapeatükis kirjeldatakse refineR algoritmi, mida kasutatakse referentsväärtuste arvutamiseks.

### 3.2.1 Wilcoxon'i astaksummatest

Wilcoxon'i astaksummatest on mitteparameetriline statistiline test, mida kasutatakse kahe sõltumatu valimi võrdlemiseks. Testi idee on moodustada valimitest ühine variatsioonirida ning seejärel võrrelda mõlema valimi astakute summasid. Nullhüpotees on, et valimid on samast jaotusest ehk nende vahel puudub statistiliselt oluline erinevus. Alternatiivne hüpotees on, et valimid ei ole samast jaotusest ning see võetakse vastu, kui p-väärtus on väiksem olulisuse nivoost ( $\alpha = 0,05$ ). Wilcoxon'i astaksummatesti eeliseks on see, et see ei eelda normaaljaotust. (Kim, 2014)

### 3.2.2 Boxi-Coxi teisendus

Boxi-Coxi teisenduse kohta antav järgnev ülevaade põhineb allikatel (Box ja Cox, 1964) ja (Ichihara jt, 2023). Boxi-Coxi teisendus on statistiline meetod andmete normaliseerimiseks. Seda kasutatakse andmete peal, mille jaotus ei järgi normaaljaotust või on kummalegi poole suure kaldega. Boxi-Coxi teisendus muudab andmete jaotuse sümmeetrilisemaks ja aitab arvutada täpsemaid referentsväärtusi. Boxi-Coxi teisenduse peamine parameeter on  $\lambda$ , mis määrab, kui palju jaotuse kuju muutub. Kui  $\lambda = 1$ , siis jääb jaotus samaks. Muutuja  $Y$  üheparameetriline Boxi-Coxi teisendus parameetri  $\lambda$  korral on kujul:

$$Y^{(\lambda)} = \begin{cases} \frac{Y^\lambda - 1}{\lambda}, & \text{kui } \lambda \neq 0, \\ \log Y, & \text{kui } \lambda = 0. \end{cases}$$

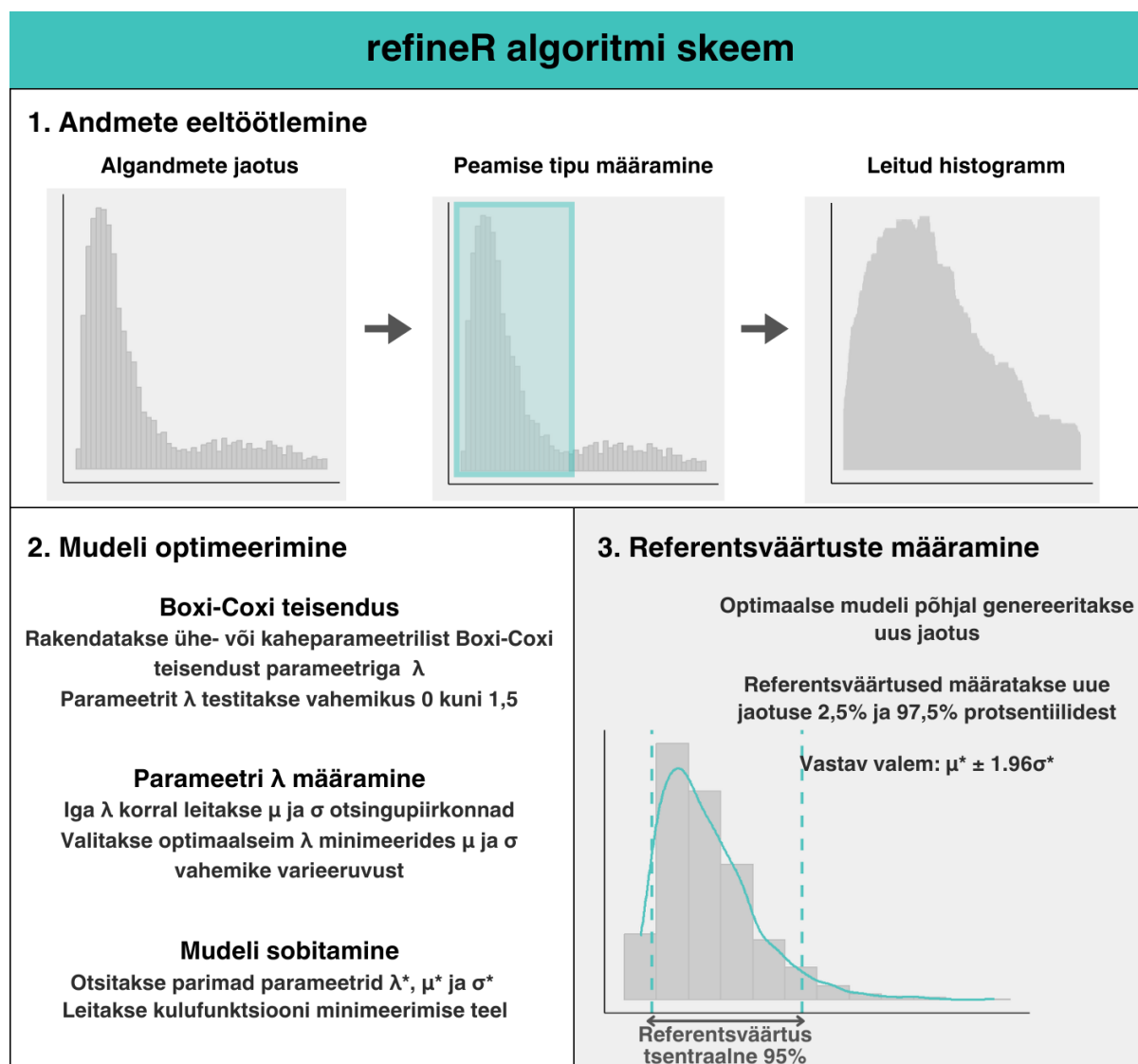
kus  $Y > 0$ . Kui andmestik sisaldab negatiivseid või nullväärtusi, kasutatakse kaheparameetrilist Boxi-Coxi teisendust, kus lisatakse nihe  $\alpha$ , et kõik väärtused oleksid positiivsed ( $Y + \alpha > 0$ ):

$$Y^{(\lambda)} = \begin{cases} \frac{(Y + \alpha)^\lambda - 1}{\lambda}, & \text{kui } (\lambda \neq 0), \\ \log(Y + \alpha), & \text{kui } (\lambda = 0). \end{cases}$$

Kui andmetel on tugev asümmeetria, siis kasutatakse kaheparameetrilist Boxi-Coxi teisendust, mis aitab andmeid ühtlustada ja paremini normaliseerida.

### 3.2.3 RefineR algoritm

Magistritöös kasutatakse referentsväärtuste arvutamiseks refineR algoritmi. Käesolev alapeatükk kirjeldab täpsemalt, kuidas algoritm töötab. Algoritmi lihtsustatud skeem on toodud joonisel 1. (Ammer jt, 2021)



Joonis 1. RefineR algoritmi tööd kirjeldav skeem.

#### Andmete eeltöötlemine

Algoritmi esimene samm on andmete puhastamine, mille käigus leitakse jaotuse peamine tipp ning eemaldatakse võimalikud erandid. Jaotuse peamine tipp valitakse suurima AUC (*Area Under the Curve*) väärtuse alusel. Antud meetodis viitab AUC histogrammi tihedusfunktsiooni alla jäävale pindalale.

Seejärel rakendab algoritm andmete normaliseerimiseks ühe- või kaheparameetrilist Boxi-Coxi teisendust parameetriga  $\lambda$ . Iga  $\lambda$  kohta vahemikus 0 kuni 1,5 määratakse otsingupiirkonnad parameetritele  $\mu$  ja  $\sigma$ . Parameeter  $\lambda$  määrab, kui palju jaotuse kuju muutub. Parameeter  $\mu$  tähistab teisendatud jaotuse keskmist ja  $\sigma$  tähistab teisendatud jaotuse standardhälvet. Parameeter  $\lambda$  määrab jaotuse kuju:

- $\lambda < 1$ : jaotuse saba on paremale;
- $\lambda = 0$ : logaritmiline normaaljaotus;
- $\lambda = 1$ : normaaljaotus;
- $\lambda > 1$ : jaotuse saba on vasakule;

Iga  $\lambda$  korral hinnatakse teisendatud andmete tihedusjaotust mitteparameetriliselt ning peamine tipp valitakse taas suurima AUC väärtuse järgi. Selle järgi määratakse parameetritele  $\mu$  parim otsingupiirkond vastavalt sammudele:

- algoritm tuvastab, kui lai on jaotuse tipp erinevatel kõrgustel;
- leitakse jaotuse keskpunkt igal kõrgusel;
- saadud keskpunktide vahemikud on parameetri  $\mu$  otsingupiirkond.

Järgmiseks määratakse parameetritele  $\sigma$  parim otsingupiirkond vastavalt sammudele:

- algoritm tuvastab, kui lai on jaotuse tipp erinevatel kõrgustel;
- igal laiusel leitakse selle standardhälve;
- saadud standardhälvete vahemikud on vastava  $\lambda$  parameetri  $\sigma$  otsingupiirkond.

Seega on iga  $\lambda$  jaoks määratud vastavad normaliseeritud  $\mu$  ja  $\sigma$  vahemikud. Seejärel valitakse kõige optimaalsem  $\lambda$  nii, et  $\mu$  ja  $\sigma$  vahemike varieerumine on minimaalne. Niimoodi saab andmed viia võimalikult sümmeetrilisele kujule ja normaaljaotuse lähedale.

### **Mudeli optimeerimine**

Algoritmi järgmine eesmärk on leida kõige optimaalsem mudel andmete kirjeldamiseks. Peale parima  $\lambda$  valimist otsitakse kõige optimaalsemad  $\mu$  ja  $\sigma$  nende eelnevalt määratletud otsingupiirkondadest. Selleks kasutatakse mitmetasandilist võrguotsingut, mille käigus arvutatakse iga parameetrite  $\lambda^*$ ,  $\mu^*$  ja  $\sigma^*$  kombinatsiooni kohta Boxi-Coxi teisendatud normaaljaotuse tihedus. Optimaalsed parameetrid leitakse kulufunktsiooni minimeerimise teel.

Kulufunktsioonil on kaks osa, millest üks maksimeerib tõenäosust, et histogrammi usaldusreale jäävaid andmeid saab seletada hinnatava mudeliga. Teine osa aitab veenduda, et tulemused on usaldusväärsed, isegi kui andmed sisaldavad ebanormaalseid tulemusi. Kulufunktsiooni väärtus näitab, kui hästi mudel andmetega sobitub. Kuna kulufunktsioon on spetsiifiline iga andmestiku jaoks, ei saa selle väärtusi erinevate andmestike vahel võrrelda. Lõpuks määrab algoritm parimad parameetrid  $\lambda^*$ ,  $\mu^*$ ,  $\sigma^*$ , mille korral kulufunktsiooni väärtus on minimaalne.

### **Referentsväärtuste hindamine**

Kui optimaalsed parameetrid on leitud, konstrueerib algoritm nende põhjal jaotuse, mis esindab kõige paremini normaalväärtusi. Referentsvahemik määratakse selle jaotuse alusel, kasutades valemit  $\mu^* \pm 1,96\sigma^*$ , mis vastab 2,5% ja 97,5% protsentilidele.

### **Usaldusintervallid *bootstrap*-meetodil**

Referentsväärtuste usaldusintervallide leidmiseks kasutatakse *bootstrap*-meetodit. Selleks valitakse  $n$  korda andmetest juhuvalim suurusega  $n$ . Iga saadud juhuvalimit kasutatakse refineR algoritmi sisendina ning leitakse iga kord vastavad referentsväärtused. Pärast  $n$  iteratsiooni leitakse usaldusintervallid, kasutades saadud alumiste ja ülemiste referentspiiride 2,5% ja 97,5% protsentiiile. See aitab paremini hinnata, kui täpselt määratud referentsväärtused kajastavad tegelikke normaalseid analüüside tulemusi ja kui suur on määramatuse ulatus. Suurus  $n$  on kasutaja poolt määratav. Kui kasutatakse *bootstrap*-meetodit, saab algoritmis määrata, et referentsväärtused määratakse kõigi saadud valimite tulemuste mediaani. See on soovitatav algoritmi autorite poolt ning seda lähenemist kasutatakse ka käesolevas töös.

Magistritöö praktilise osa analüüsiks on kasutatud statistikatarkvara R (versioonid 4.4.0 ja 4.3.2). Töö vormistati tekstitöötlusprogrammi  $\text{\LaTeX}$  abiga. Tulemuste visualiseerimiseks on jooniste tegemiseks kasutatud R ja Shiny R keskkonda. Kogu töö visuaalse ülevaate ja algoritmi skeemi tegemiseks on kasutatud graafilise disaini tööriista Canva.

### **3.2.4 Eetikakomitee luba**

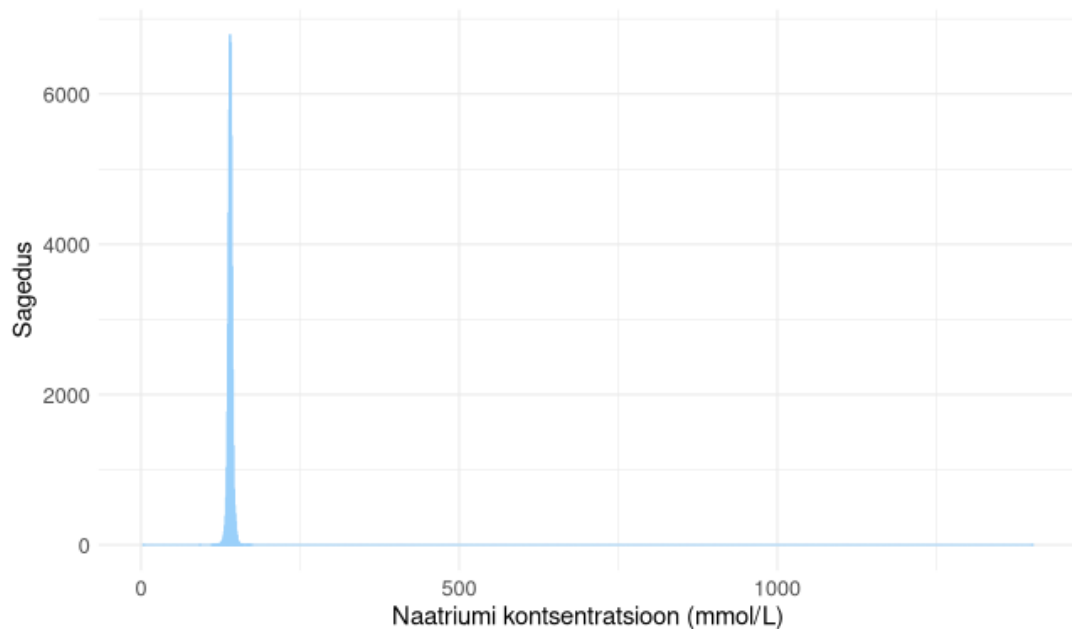
Käesolev uuring on läbi viidud vastavalt Eesti bioetika ja inimuuringute nõukogu loale nr 1.1-12/611.

## 4. Tulemused

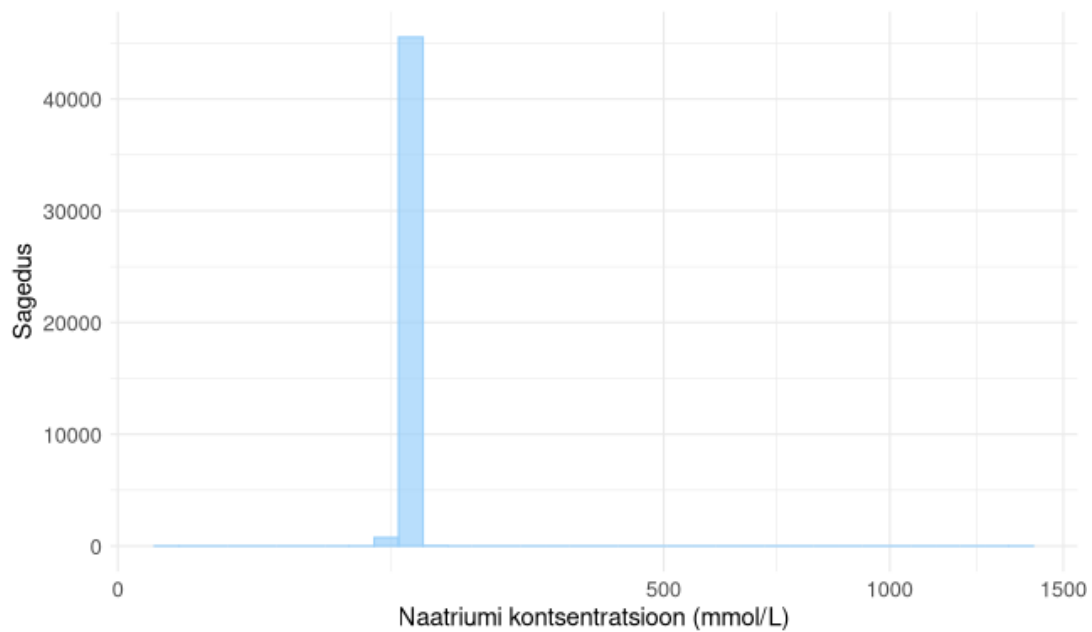
Käesolev peatükk annab ülevaate töö tulemustest.

### 4.1 Naatrium

Naatriumi analüüside tulemuste jaotus on kujutatud joonisel 2. Jooniselt paistab, et jaotus on pikema parema sabaga, kuid peamine tipp, kuhu enamus andmeid on kogunenud, on sümmeetriline. Joonisel 3 on kujutatud sama jaotust, kus x-telg on esitatud ruutjuure skaalal. Selline teisendus aitab väiksemad väärtused paremini nähtavaks teha ja vähendab suurte väärtuste mõju jaotuse üldisele kujule. Ka sellelt jooniselt on näha, et enamus andmed paiknevad ühes piirkonnas ning paremal esinevad üksikud erandid.

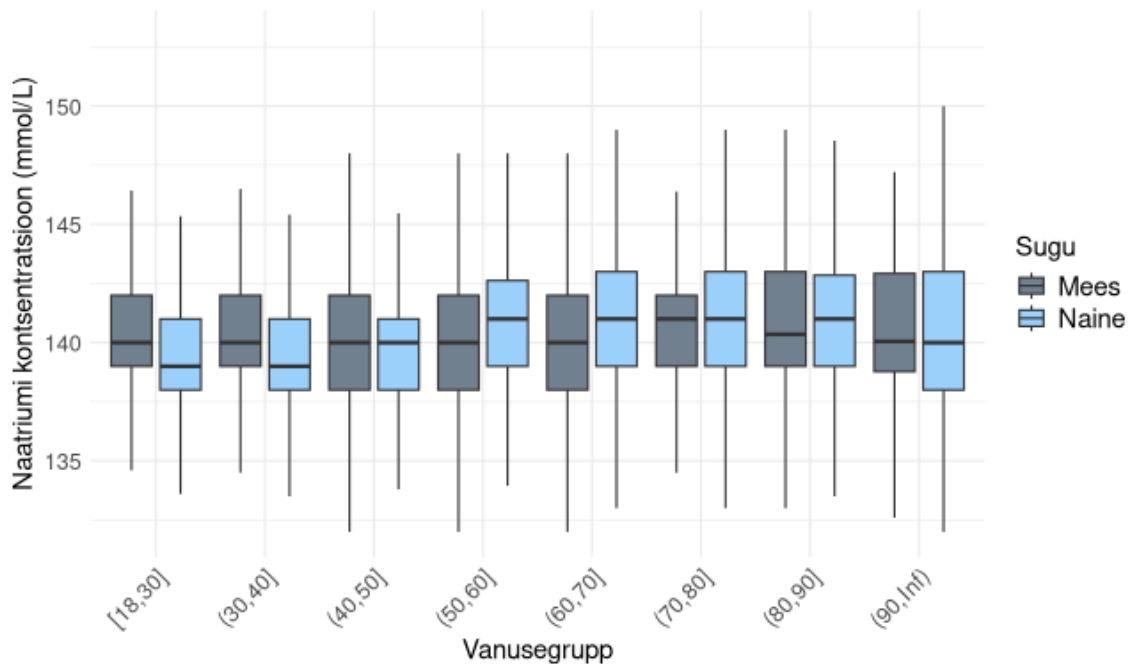


Joonis 2. Naatriumi analüüsitulemuste jaotus.



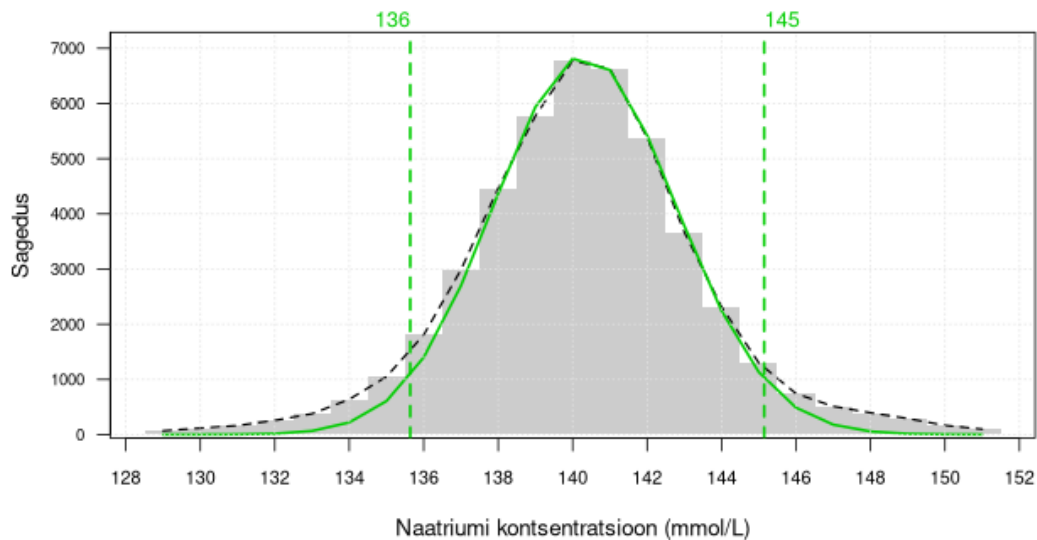
Joonis 3. Naatriumi analüüsitulemuste jaotus, kus x-telg on esitatud ruutjuure skaalal.

Joonisel 4 on kujutatud naatriumi analüüside tulemuste jaotust vanuse ja soo lõikes. Ka sealt paistab, et naatriumi sisaldus veres püsib suhteliselt ühtlasena igas vanusegrupis ning ka meeste ja naiste tulemused ei erine oluliselt. Wilcoxon'i astaksummatest ei näidanud meeste ja naiste analüüsitulemuste vahel statistiliselt olulist erinevust ( $p = 0,19$ ).



Joonis 4. Naatriumi analüüsitulemused vanuse ja soo lõikes.

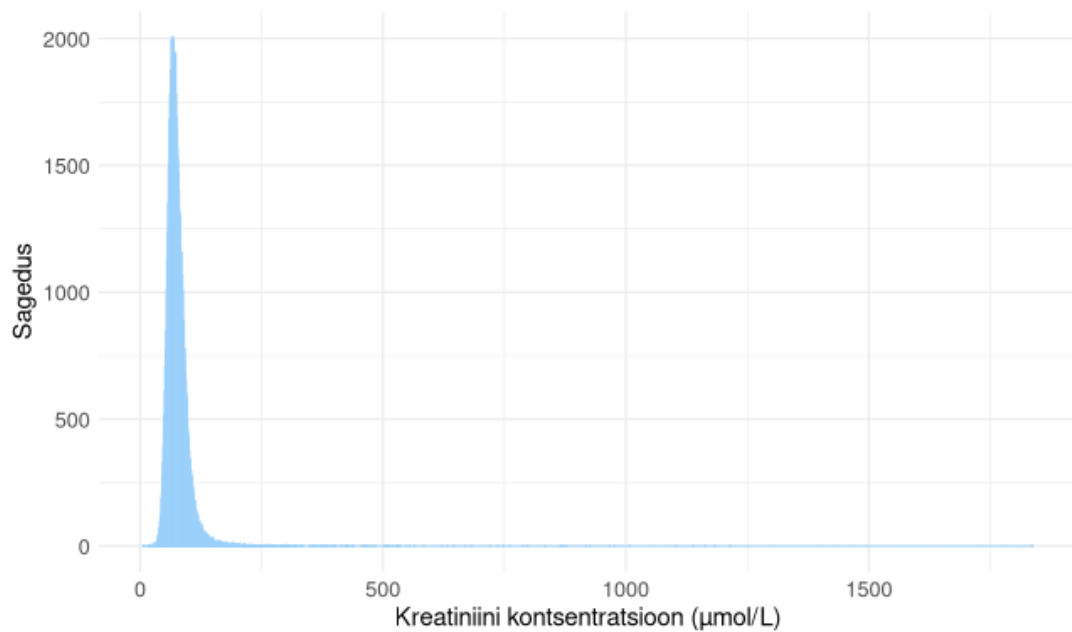
Seetõttu määratakse naatriumile üks üldine referentsväärtus vähemalt 18-aastaste inimeste kohta. RefineR algoritmis kasutatakse üheparameetrilist Boxi-Coxi teisendust koos usaldusintervallidega. Leitud referentsväärtused koos 95% usaldusintervallidega on vahemikus 136 (135; 136)–145 (145; 146) mmol/L (vt. joonis 5). Kõigi analüütide täpsemad referentsväärtused koos usaldusintervallidega on esitatud lisas I olevas tabelis 3.



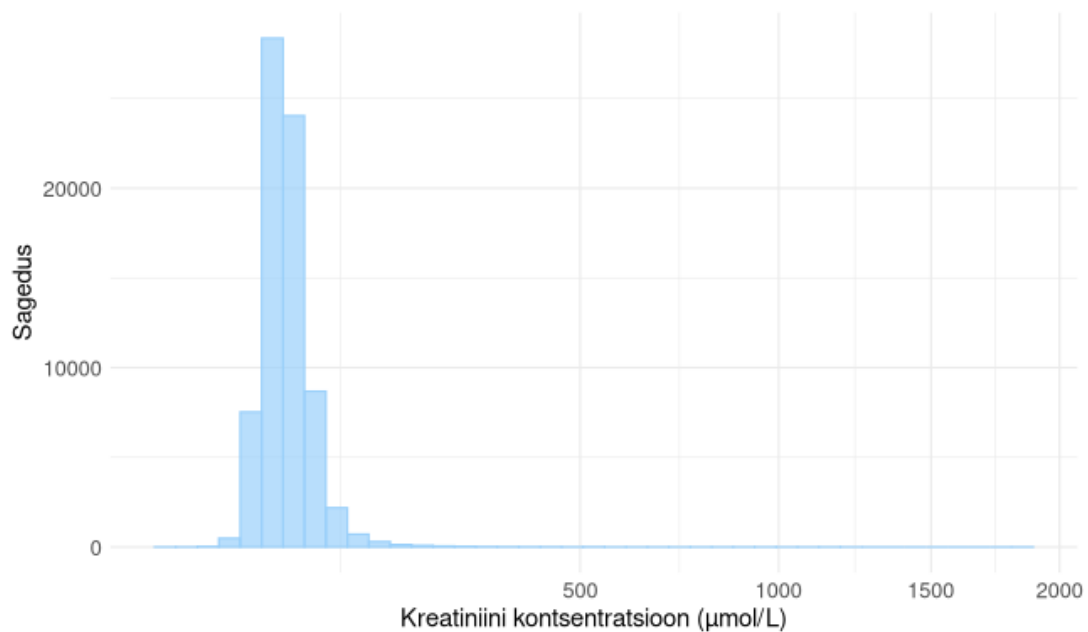
Joonis 5. Naatriumi referentsväärtused.

## 4.2 Kreatiniin

Kreatiniini analüüside tulemuste jaotus ei ole sümmeetriline - jaotus on paremale kaldu ning pika peenikese sabaga (vt joonis 6). Samas on suurem osa andmeid koondunud ühe selge tipu ümber, mis on sümmeetrilisema kujuga. Joonisel 7 on sama jaotus esitatud ruutjuure skaalal, mis aitab paremini esile tuua põhilise osa andmetest. Sellisel kujul paistab jaotus visuaalselt normaaljaotuse sarnasena, kuigi jääb endiselt paremale kaldu.

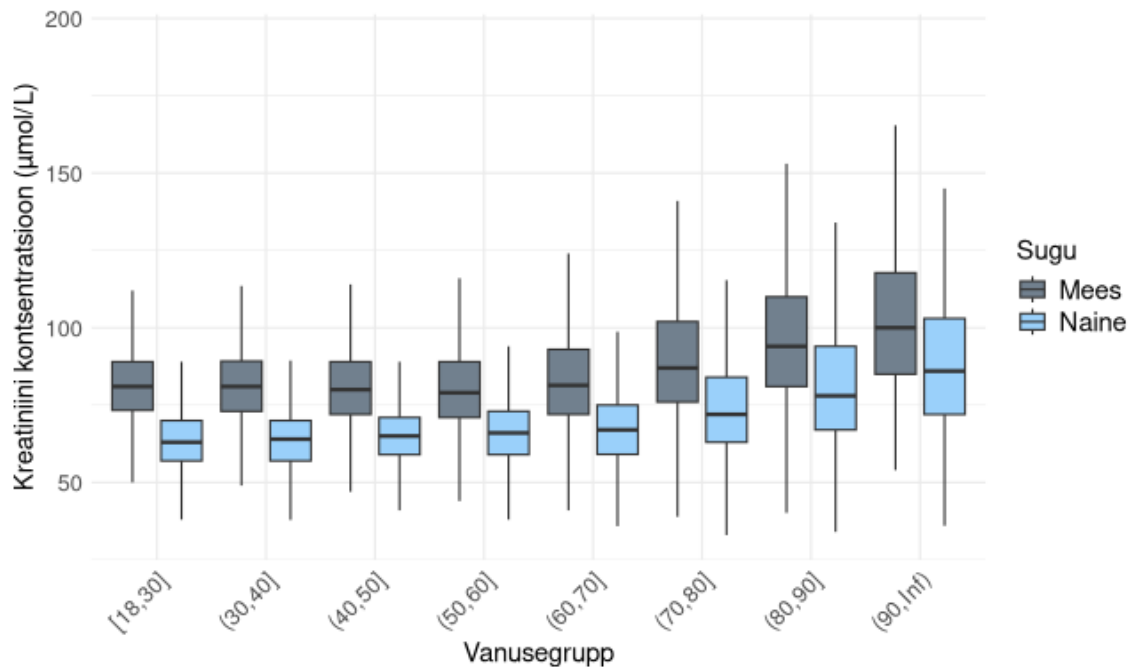


Joonis 6. Kreatiniini analüüsitulemuste jaotus.



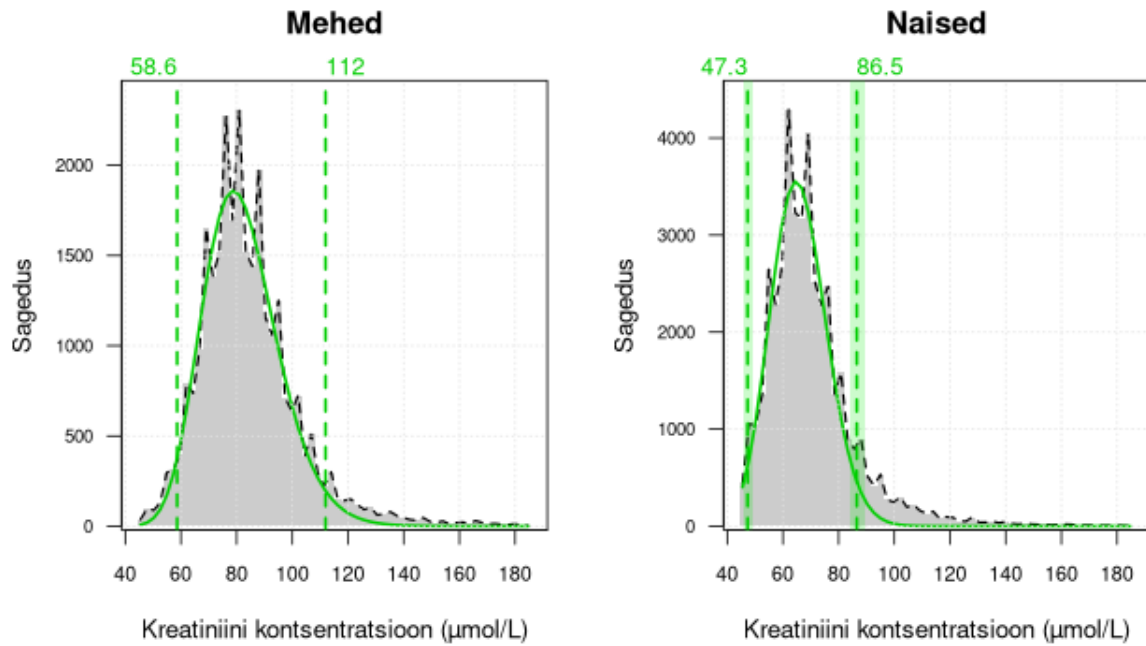
Joonis 7. Kreatiniini analüüsitulemuste jaotus. X-telg on esitatud ruutjuure skaalal.

Joonisel 8 on kujutatud kreatiniini jaotus vanusegruppide ja soo lõikes, kus paistab, et meeste tulemused on keskmiselt kõrgemad kui naistel. Wilcoxon'i astaksummatesti tulemuste põhjal ilmneb meeste ja naiste kreatiniini analüüsitulemuste vahel statistiliselt oluline erinevus ( $p < 0,05$ ).



Joonis 8. Kreatiniini analüüsitulemused vanuse ja soo lõikes.

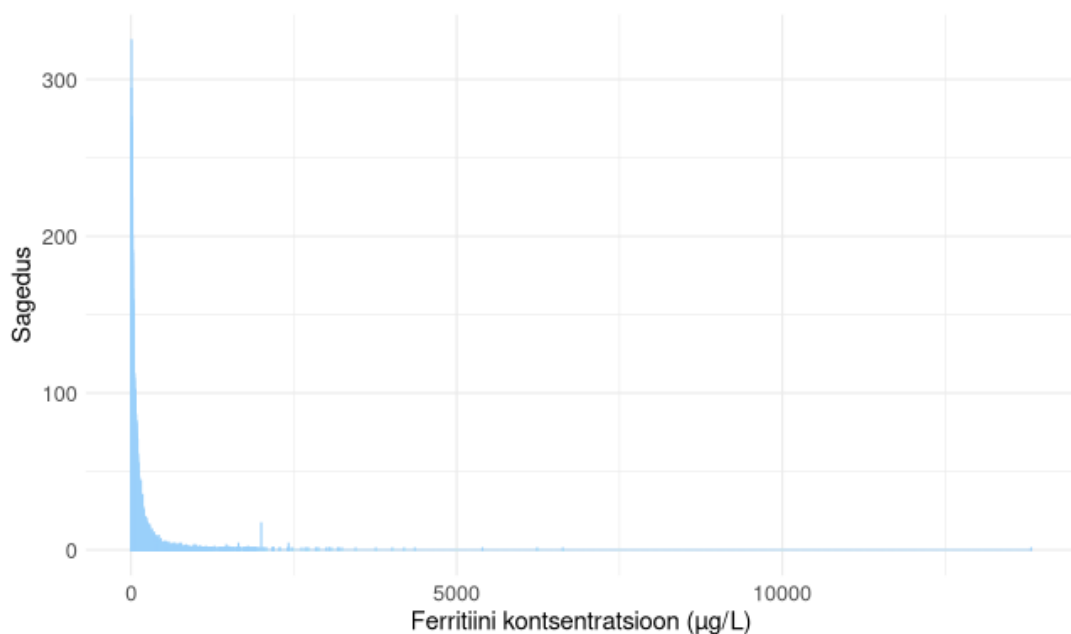
Kuna meeste ja naiste tulemused on statistiliselt erinevad, siis määratakse kreatiniinile soospetsiifilised referentsväärtused. Selleks kasutatakse üheparameetrilist Boxi-Coxi teisendust koos usaldusintervallidega. Leitud kreatiniini referentsvahemikud koos usaldusintervallidega on meestel 59 (57; 59)–112 (106; 113)  $\mu\text{mol/L}$  ning naistel 47 (46; 49)–87 (84; 89)  $\mu\text{mol/L}$  (vt. joonis 9).



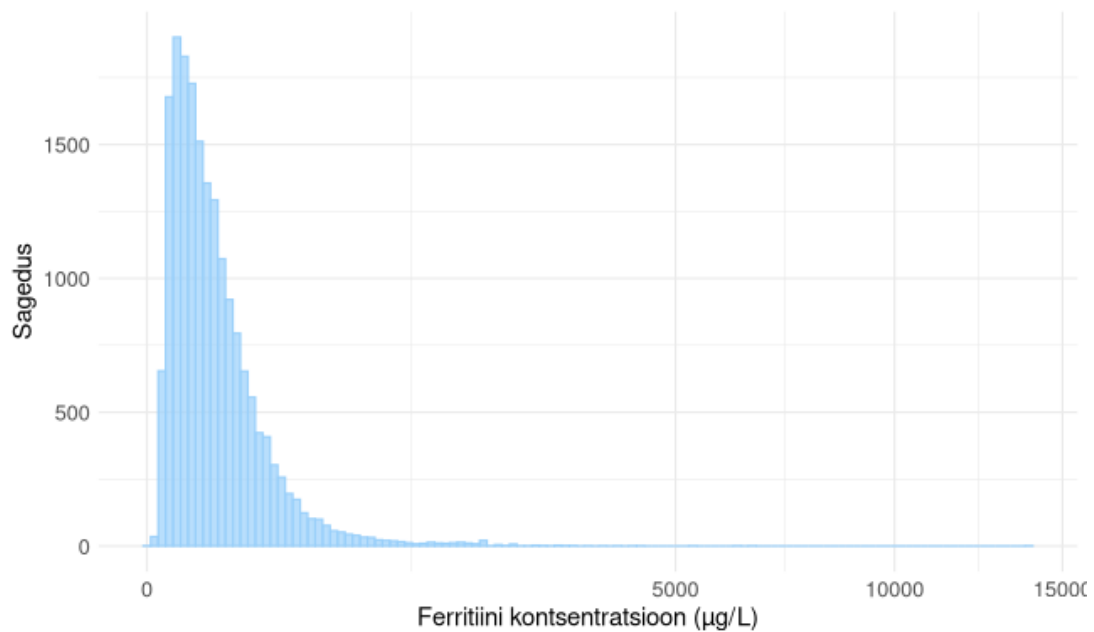
Joonis 9. Kreatiniini referentsväärtused koos 95% usaldusintervallidega.

### 4.3 Ferritiin

Ferritiini analüüsitulemuste jaotus on kujutatud joonisel 10. Selgelt paistab, et tulemuste jaotus on tugeva parema kaldega ja sabaga. Suurem osa andmeid on koondunud ühe selge tipu ümber, aga ka selle jaotus on paremale kaldu. Joonisel 11 on kujutatud sama jaotust ruutjuure skaalal. Sellelt jooniselt paistab selgemalt, et ferritiini tulemuste jaotus on parema kaldega.

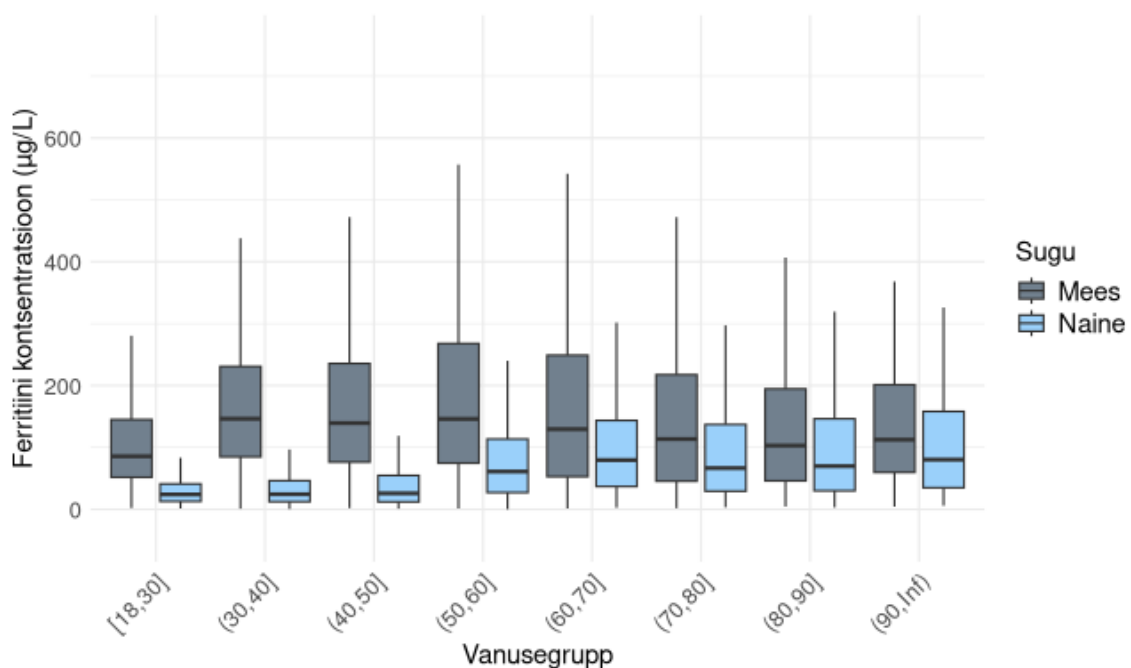


Joonis 10. Ferritiini analüüsitulemuste jaotus.



Joonis 11. Ferritiini analüüsitulemuste jaotus, kus x-telg on esitatud ruutjuure skaalal.

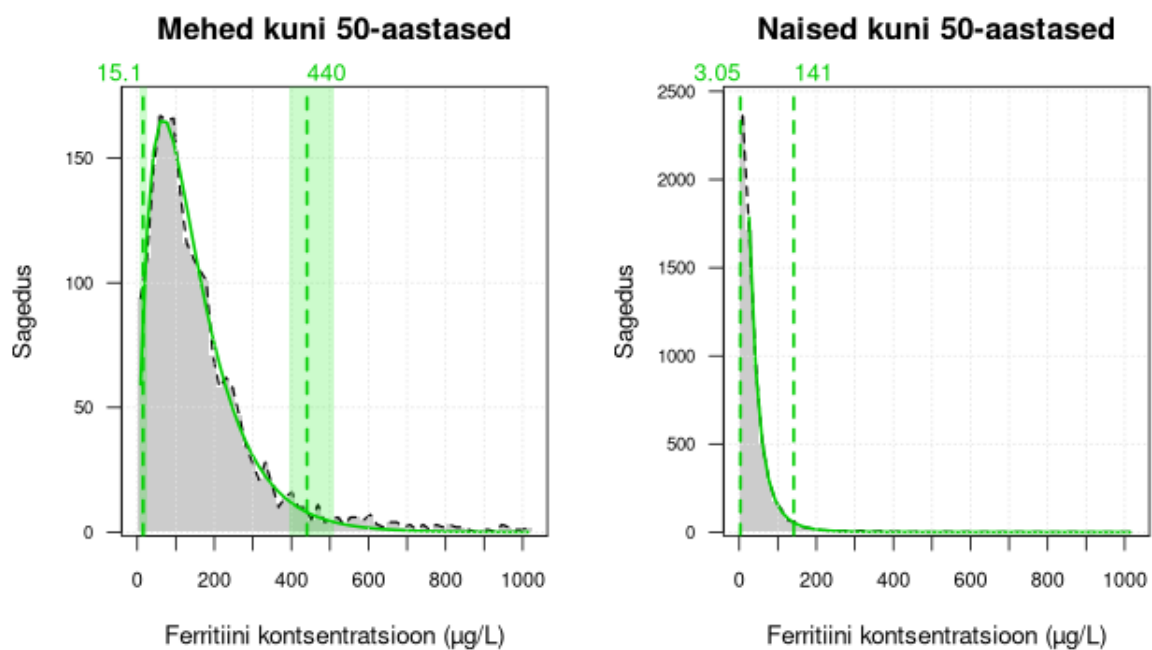
Ferritiini analüüside tulemused esitatakse samuti ka eraldi soo ja vanuserühmade kaupa. Joonisel 12 on näha, et meeste ferritiini tase on kõikides vanusegruppides kõrgem kui naistel. Wilcoxon'i astaksummatest näitab, et meeste ja naiste ferritiini analüüsitulemuste vahel on statistiliselt oluline erinevus ( $p < 0,05$ ). Naiste puhul on märgatav ferritiini taseme tõus alates 50. eluaastast, mis viitab vanusega seotud muutustele.



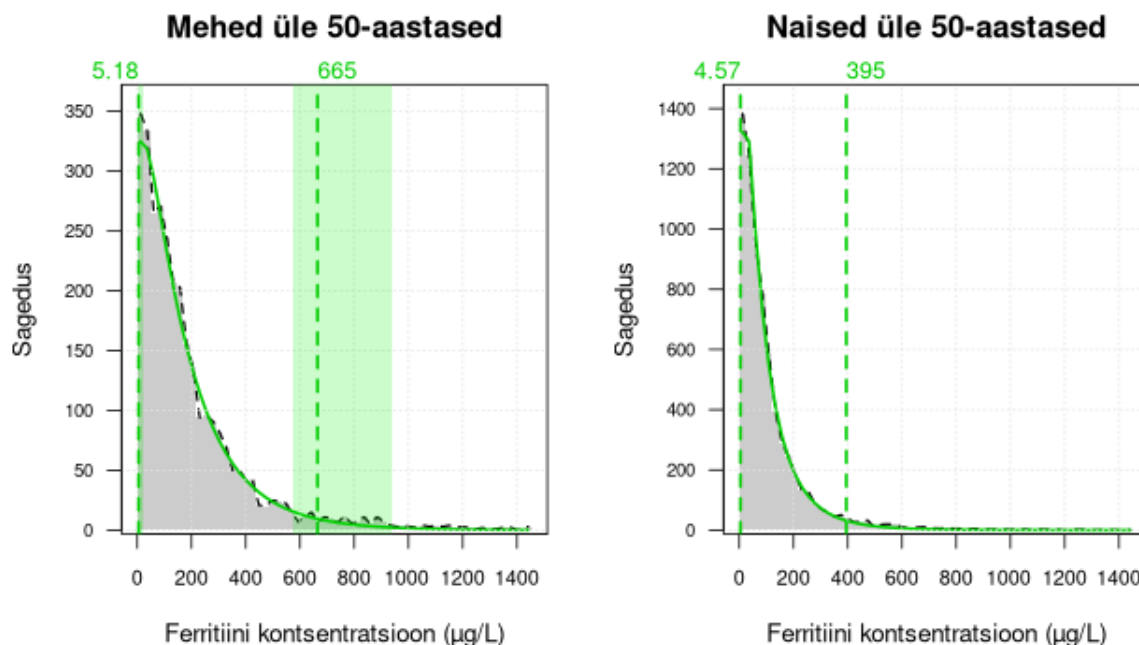
Joonis 12. Ferritiini analüüsitulemused vanuse ja soo lõikes.

Seetõttu määratakse ferritiini referentsväärtused soo ja vanuse põhjal. Kasutatakse refineR algoritmi koos kaheparameetrilise Boxi-Coxi teisendusega koos usaldusintervallidega. Saadud ferritiini referentsvahemikud koos 95% usaldusintervallidega on:

- meestel (kuni 50-aastased): 15 (8; 24)–440 (392; 511)  $\mu\text{g/L}$  (vt. joonis 13);
- naistel (kuni 50-aastased): 3 (2; 4)–141 (108; 157)  $\mu\text{g/L}$  (vt. joonis 13);
- meestel (üle 50-aastased): 5 (2; 25)–665 (572; 942)  $\mu\text{g/L}$  (vt. joonis 14);
- naistel (üle 50-aastased): 5 (2; 12)–395 (297; 543)  $\mu\text{g/L}$  (vt. joonis 14).



Joonis 13. Ferritiini referentsväärtused koos 95% usaldusintervallidega kuni 50-aastastel.



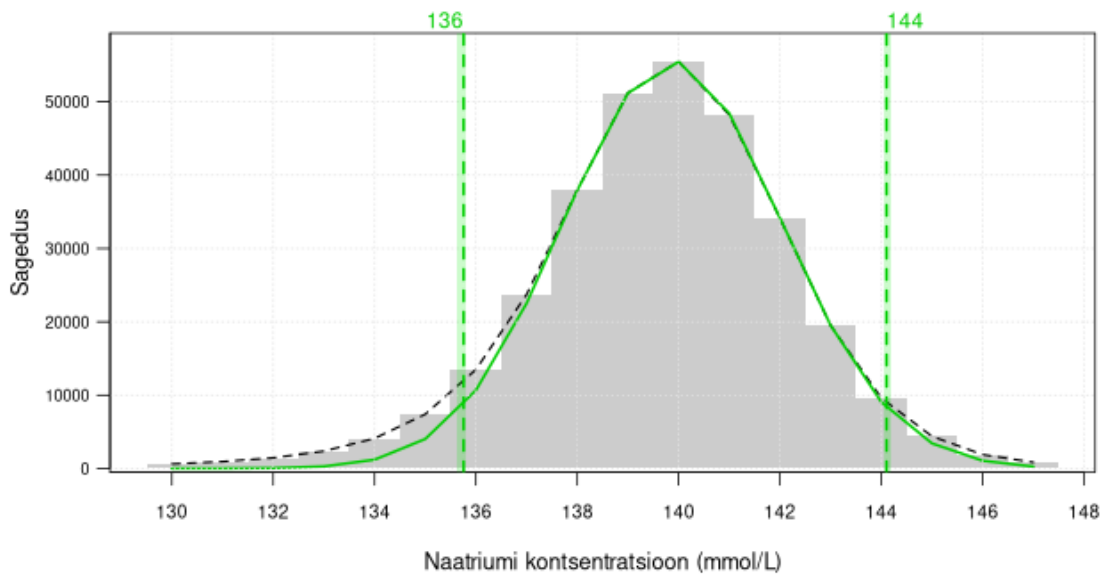
Joonis 14. Ferritiini referentsväärtused koos 95% usaldusintervallidega üle 50-aastastel.

#### 4.4 Lisa-analüüs tervise infosüsteemi andmetel

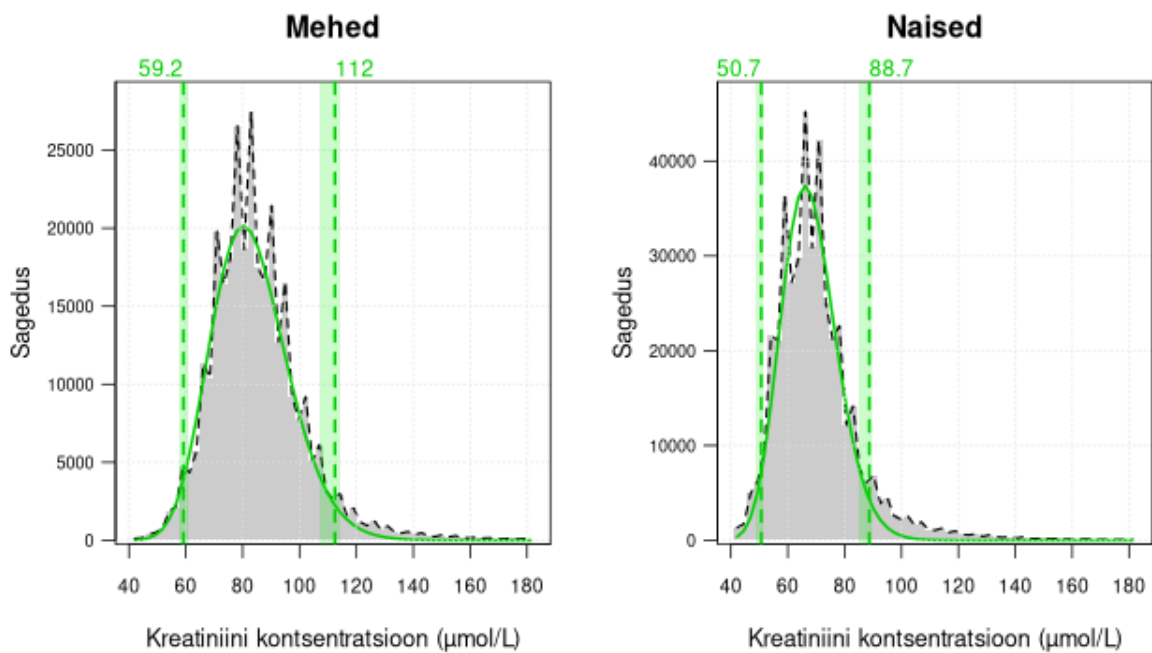
Välja töötatud meetodikat prooviti lisaks ka uuemate andmete peal, et hinnata selle rakendatavust ja tulemuste stabiilsust. Lisa-analüüs viidi läbi tervise infosüsteemi 2023. kuni 2024. aasta andmetel ning keskendus samadele kolmele analüüdile – naatrium, kreatiniin ja ferritiin. Referentsväärtuste määramisel kasutatakse sama lähenemist, mida eelnevalt kasutati. See tähendab, et leitakse samade parameetrite ja sama soo- ja vanuse-spetsiifilised väärtused.

Jooniselt 15 on näha tervise infosüsteemi andmestiku põhjal määratud naatriumi referentsväärtused koos 95% usaldusintervallidega, vahemikuks määrati 136 (136; 136)–144 (144; 144) mmol/L. Kreatiniini referentsvahemikuks määrati meestel 59 (58; 61)–112 (107; 114)  $\mu\text{mol/L}$  ning naistel 51 (49; 51)–89 (85; 89)  $\mu\text{mol/L}$  (vt. joonis 16). Ferritiini analüüdile määratud referentsvahemikud koos 95% usaldusintervallidega on:

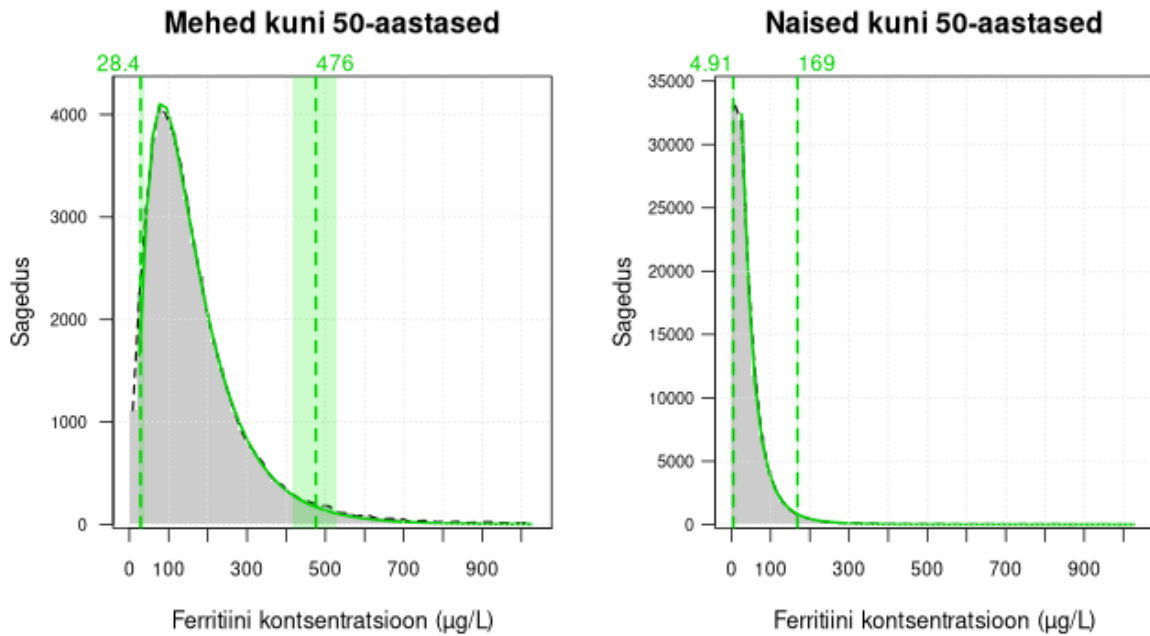
- meestel (kuni 50-aastased): 28 (23; 39)–476 (418; 528)  $\mu\text{g/L}$  (vt. joonis 17);
- naistel (kuni 50-aastased): 5 (4; 5)–169 (165; 173)  $\mu\text{g/L}$  (vt. joonis 17);
- meestel (üle 50-aastased): 16 (7; 29)–683 (567; 796)  $\mu\text{g/L}$  (vt. joonis 18);
- naistel (üle 50-aastased): 15 (7; 20)–372 (325; 434)  $\mu\text{g/L}$  (vt. joonis 18).



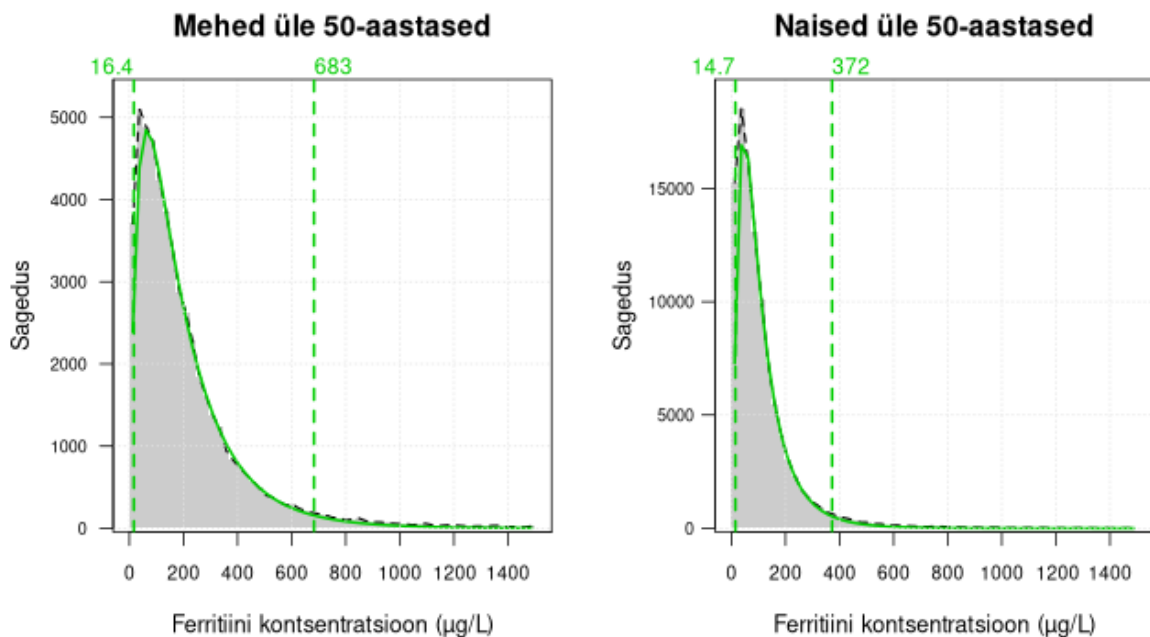
Joonis 15. Uute andmete põhjal määratud naatriumi referentsväärtused koos usaldusintervallidega.



Joonis 16. Uute andmete põhjal määratud kreatiniini referentsväärtused koos 95% usaldusintervallidega.



Joonis 17. Uute andmete põhjal määratud ferritiini referentsväärtused koos 95% usaldusintervallidega kuni 50-aastastel.



Joonis 18. Uute andmete põhjal määratud ferritiini referentsväärtused koos 95% usaldusintervallidega üle 50-aastastel.

Lisa-analüüsi käigus määratud referentsväärtused koos usaldusintervallidega on esitatud lisa I olevas tabelis 4.

## 4.5 Kasutajaliides

Referentsväärtuste arvutamise ja visualiseerimise lihtsustamiseks on loodud R Shiny keskkonnas kasutajaliides. See võimaldab kasutajatel üles laadida oma andmeid, neid töödelda, analüüsida ning visualiseerida tulemusi. Liides muudab referentsväärtuste määramise lihtsamaks ja kättesaadavamaks.

Kasutajaliides on jaotatud neljaks vaheleheks. Esimesel vahelehel saab kasutaja ülevaate rakenduse kasutusvõimalustest ning seejärel on võimalik andmed üles laadida CSV või Exceli formaadis (vt. lisa II). Seejärel tuleb määrata, kas failil on päis, millist veergude eraldajat kasutatakse, analüüdi nimetus ning millised veerud sisaldavad pseudonüümitud patsiendi identifikaatorit, sugu, vanust ja analüüsitulemusi. Lisaks tuleb täpsustada, kuidas on andmetes tähistatud mees- ja naissoost isikud. Pärast andmete üleslaadimist eemaldab liides automaatselt sama patsiendi korduvad mõõtmised, mis on tehtud ühe aasta jooksul ning tühjad väärtused.

Teisel vahelehel on ülevaade andmestikust ning interaktiivne histogramm analüüsitulemuste jaotusest (vt. lisa III). See võimaldab hinnata andmete jaotust ja võimalikke kaldeid, mis omakorda aitab otsustada, kas referentsväärtuste määramisel kasutada ühe- või kaheparameetrilist Boxi-Coxi teisendust. Näiteks tugeva kalde korral tuleks kasutada kaheparameetrilist teisendust.

Kolmandal vahelehel on karpdiagramm, mis näitab tulemuste jaotust soo ja vanusegruppide lõikes (vt. lisa IV). See aitab hinnata, kas sugu ja vanus mõjutavad tulemusi ning kas referentsväärtused tuleks määrata vanuserühmadele eraldi.

Viimasel vahelehel määratakse analüüdile referentsväärtused (vt. lisa V). Kõige pealt kuvatakse Wilcoxon'i astaksummetesti tulemus ning selle põhjal soovitus, kas määrata soospetsiifilised väärtused. Kui erinevus meeste ja naiste analüüsitulemuste vahel on statistiliselt ja kliiniliselt oluline, siis on soovitatav määrata soospetsiifilised referentsväärtused. Kasutaja saab valida parameetrid, mille alusel määratakse referentsväärtused refineR algoritmi abil. Saadud väärtusi visualiseeritakse joonisel. Joonisel on näha visuaalselt võrdlust tegelike ja mudeli jaotuse vahel ning referentsväärtused vajadusel koos usaldusintervallidega. See võimaldab kasutajal visuaalselt hinnata, kui hästi leitud mudel sobitub tegelike andmetega.

Kasutajaliidese eesmärk on muuta analüüside tulemuste analüüs ja referentsväärtuste määramine kiiremaks ja kasutajasõbralikumaks. Loodud kasutajaliidese ja analüüsi koodid on saadaval GitLabi projektis, mille link on lisas VI.

## 5. Arutelu

Magistritöös analüüsi naatriumi, kreatiniini ja ferritiini analüüside tulemusi ning määrati nendele analüütidele referentsväärtused Eesti rahvastiku näitel. Selles peatükis arutletakse peatükis 4 esitatud tulemuste üle ning võrreldakse töös määratud referentsväärtusi alapeatükis 2.3 tabelis 1 välja toodud hetkel kasutusel olevate väärtustega.

### 5.1 Naatrium

Naatriumi referentsväärtuste leidmisel selgus, et selle tase veres püsib üsna stabiilsena nii meeste kui ka naiste ning erinevate vanusegruppide lõikes. Analüüside tulemuste jaotus ei näidanud erinevusi soo ega vanuse lõikes. Wilcoxon'i astaksummatest näitas, et statistiliselt olulist erinevust meeste ja naiste tulemuste vahel ei ole ( $p = 0,19$ ). Mistõttu määrati naatriumile üks üldine referentsväärtus alates 18. eluaastast. Tabelis 2 võimaldab võrrelda rahvusvaheliselt ja Eesti laborites kasutatavaid referentsväärtusi käesolevas töös leitud väärtustega.

Tabel 2. Kolme analüüdi referentsväärtuste võrdlus

	<b>Naatrium (mmol/L)</b>	<b>Kreatiniin (<math>\mu\text{mol/L}</math>)</b>	<b>Ferritiin (<math>\mu\text{g/L}</math>)</b>
<b>Kanada labor</b>	$\geq 18$ a 136–146	$\geq 18$ a 49–93 (M) $\geq 18$ a 22–75 (N)	$\geq 18$ a 24–336 (M) $\geq 18$ a 11–307 (N)
<b>TÜK Ühendlabor</b>	$\geq 18$ a 136–145	$\geq 15$ a 59–104 (M) $\geq 15$ a 45–84 (N)	18–61 a 30–400 (M) 18–61 a 13–150 (N)
<b>SYNLAB</b>	$\geq 19$ a 136–145	$\geq 19$ a 64–111 (M) $\geq 19$ a 50–98 (N)	$\geq 19$ a 28–370 (M) $\geq 19$ a 10–204 (N)
<b>RITA MAITT andmetel</b>	$\geq 18$ a 136–145	$\geq 18$ a 59–112 (M) $\geq 18$ a 47–87 (N)	18–50 a 15–440 (M) 18–50 a 3–141 (N) $\geq 51$ a 5–665 (M) $\geq 51$ a 5–395 (N)
<b>Tervise infosüsteemi andmetel</b>	$\geq 18$ a 136–144	$\geq 18$ a 59–112 (M) $\geq 18$ a 51–89 (N)	18–50 a 28–476 (M) 18–50 a 5–169 (N) $\geq 51$ a 16–683 (M) $\geq 51$ a 15–372 (N)

Kuna naatriumi tulemuste jaotus oli suhteliselt sümmeetriline ega kaldunud tugevalt kõrvale normaaljaotusest (vt. joonis 2), siis peeti üheparameetrilist Boxi-Coxi teisendust sobivaks. Täpsemate tulemuste saamiseks, rakendati algoritmi koos *bootstrap*-meetodiga, kus  $n = 100$ .

Rahvusvaheliselt kasutatakse samuti naatriumi referentsväärtusi, mis ei ole soospetsiifilised, kuid võivad mõnevõrra erineda sõltuvalt laborist. Tabelist 2 on näha, et töös saadud naatriumi referentsväärtused kattuvad täielikult Eesti laborite omadega. Kanadas kasutusel olevas referentsväärtuses erineb ainult ülemine referentsväärtus ühe ühiku võrra. Magistritöös leitud referentsvahemikust 136–145 mmol/L võib lugeda seega usaldusväärseks ja see kinnitab ka kasutatud meetodi usaldusväärsust. Lisaks viidi metoodika kontrollimiseks läbi lisa-analüüs tervise infosüsteemi andmetel, mille põhjal määratud naatriumi referentsväärtuse vahemik oli 136–144 mmol/L. Kuigi ülemine piir on ühiku võrra madalam, viitab see üldiselt metoodika stabiilsusele ja kinnitab leitud tulemust.

## 5.2 Kreatiniin

Joonis 6 näitas selget kreatiniini analüüsi tulemuste jaotuse erinevust soo lõikes: meestel oli kreatiniini tase oluliselt kõrgem kui naistel. Seda kinnitas ka tehtud Wilcoxon'i astaksummatest, kus  $p < 0,05$ . Seetõttu määrati kreatiniinile soospetsiifilised referentsväärtused.

RefineR algoritmi kasutamisel rakendati üheparameetrilist Boxi-Coxi teisendust, mis on sobilik arvestades jaotuse kuju. See tähendab, et andmete jaotuse peamine tipp ei ole tugeva kaldega. Stabiilsemate tulemuste saamiseks kasutati ka *bootstrap*-meetodit 100 kordusvalimiga, mis aitas määrata täpsemad referentsväärtused ja usaldusintervallid.

Magistritöö käigus määratud kreatiniini referentsvahemik on meestel 59–112  $\mu\text{mol/L}$  ja naistel 47–87  $\mu\text{mol/L}$ . Meeste kreatiniini referentsväärtused sarnanevad enim SYNLABi väärtustega ning naistel Ühendlabori väärtustega (vt. tabel 2). Lisa-analüüsis määratud referentsväärtused on töö põhianalüüsi tulemustega väga sarnased. Täpsemalt meeste puhul referentsväärtused ühtivad täielikult ja naistel on veidi kõrgemad väärtused. Töös määratud väärtused on üldiselt kooskõlas hetkel kasutusel olevate referentsväärtustega, kuid esineb väiksemaid erinevusi.

## 5.3 Ferritiin

Ferritiini tulemuste analüüs näitas olulist erinevust nii soo kui ka vanuse lõikes. Statistilist erinevust soo lõikes kinnitas tehtud Wilcoxon'i astaksummatest, kus  $p < 0,05$ . Meeste puhul püsis ferritiini tase üsna stabiilsena kõikides vanusegruppides, kuid oli keskmiselt kõrgem kui naistel. Naistel täheldati selget tõusu alates 50. eluaastast. See on kirjanduse põhjal seotud

menopausijärgse muutustega raua ainevahetuses (Jian jt, 2009). Seepärast leiti meestele ja naistele vanuse põhjal kaks referentsväärtust – kuni 50-aastaste referentsväärtused ja üle 50-aastaste referentsväärtused.

Jooniselt 8 selgus, et ferritiini analüüside tulemuste jaotus oli tugevalt paremale kaldu ning seetõttu kasutati referentsväärtuste arvutamiseks refineR algoritmi koos kaheparameetrilise Boxi-Coxi teisendusega. Lisaks rakendati bootstrapi-meetodit ( $n = 100$ ), et saada täpsemad tulemused. Referentsväärtuste laiad usaldusintervallid viitavad sellele, et analüüsitulemused varieeruvad andmestikus üsna palju. Seepärast on oluline hinnata ka metoodika stabiilsust.

Töö käigus määratud ferritiini referentsvahemik on kuni 50-aastaste meeste puhul 15–440  $\mu\text{g/L}$  ja naistel 3–141  $\mu\text{g/L}$ . Üle 50-aastaste meeste puhul määrati ferritiini referentsvahemikuks 5–665  $\mu\text{g/L}$  ja naistel 5–395  $\mu\text{g/L}$ . Tabelist 2 on näha, et saadud tulemused on üldjoontes kooskõlas varasemate laborite poolt kasutatavate referentsväärtustega, ent võimaldavad täpsemat eristamist soo ja vanuse lõikes. Töös määratud meeste ferritiini referentsväärtused on laiema vahemikuga võrreldes teiste laborite omadega. Samas naistel on saadud referentsväärtuse ülemine piir madalam kui hetkel kasutusel olevatel väärtustel. Kõik laborid ei erista ferritiini referentsväärtusi vanuse järgi või neil puuduvad üldse vanemate inimeste väärtused, mistõttu pakub antud töö väärtuslikku täiendust senistele referentsväärtustele.

Lisa-analüüsi tulemused olid üldiselt sarnased töö tulemustega, kuid esinevad ka mõned erinevused. Enamik ferritiini referentsväärtusi tulid lisa-analüüsis veidi kõrgemad. Erandiks on üle 50-aastaste naiste ülemine referentsväärtus, mis tuli veidi madalam. Tulemused näitavad vajadust testida referentsväärtuste määramise metoodikat erinevate andmestike peal, et tagada selle usaldusväärsus ja üldistatavus.

## 5.4 Kasutajaliides

Magistritöö käigus loodud kasutajaliides refineR algoritmile, koos oma mõõtmisandmete üles laadimise võimalusega, võimaldab teha andmete eeltöötlust, analüüsi ning referentsväärtuste määramist lihtsal ja kasutajasõbralikul viisil. Liidese eesmärk oli luua tööriist, mis muudab statistiliste arvutuste rakendamise kättesaadavamaks meditsiinitöötajatele ja laborianalüütikutele, kellel on vajadus määrata oma andmete põhjal uusi või kohandatud referentsväärtusi.

Liideses kasutatavad joonised toetavad kasutajat andmete mõistmisel ja otsuste tegemisel, näiteks kas määrata soospetsiifilisi väärtusi või millist Boxi-Coxi teisendust rakendada. Samuti aitab liidese loogiline ülesehitus hoida analüüsiprotsessi selgena ja sujuvana.

Töö raames valminud kasutajaliides on praktiline, sest see ühendab andmete eeltötluse, statistilise analüüsi ja visualiseerimise üheks sujuvaks töökäiguks. See vähendab vajadust käsitsi tehtavate analüüside järele ning toetab referentsväärtuste määramist erinevate parameetrite ja andmestike põhjal. Samuti laiendab see refineR algoritmi praktilist kasutust.

Kuigi loodud kasutajaliides lihtsustab referentsväärtuste määramist, ei pruugi see siiski tuvastada kõiki andmevigu. Lisaks puudub hetkel võimalus andmeid puhastada mõne teguri alusel, näiteks haiguste alusel. Tulevikus võiks liidest täiendada andmete kontrollimise ja puhastamise osas ning laiendada valikuid nii andmepuhastuses kui ka referentsväärtuste määramise juures. Need täiendused muudaksid kasutajaliidese veelgi kasutajasõbralikumaks ja aitaksid saavutada täpsemaid tulemusi.

## **5.5 Töö tugevused, nõrkused ja edasised sammud**

Magistritöö käigus määrati referentsväärtused olemasolevate terviseandmete põhjal, kasutades selleks refineR algoritmi. Sellisel lähenemisel on mitmeid tugevaid külgi. Esiteks töötas kasutatud meetod antud andmetel küllaltki hästi, kuna leitud referentsväärtused langevad üldjoontes kokku nii rahvusvaheliste kui ka Eestis kasutusel olevate väärtustega. Teiseks loodi töö käigus praktilist kasutust toetav interaktiivne kasutajaliides, mis muudab keerulisema analüüsi ligipääsetavamaks ja taaskasutatavaks ka teistele kasutajatele.

Samas on meetodil ka mitmeid piiranguid. Praegune loogika selle kohta, kas ja kuidas määrata referentsväärtused vanuse alusel eraldi, on osaliselt subjektiivne ning tugineb visuaalsele hinnangule. Tulevikus tuleks kaaluda objektiivsemat lähenemist – näiteks kasutada statistiliste testide rakendamist vanusegruppide vaheliste erinevuste hindamiseks. Samas tuleb ka arvestada, et mis on kliiniliselt oluline ja vajalik. Teise võimalusena võiks hinnata, kui suur osa inimesi jääks igas vanusegrupis määratud üldistest referentsväärtustest välja. Kui väljajääjate osakaal ületab teatud kriitilist piiri, näiteks 15–20%, võib see viidata vajadusele määrata vanusespetsiifilised referentsväärtused. Sellised lähenemised aitaksid otsustusprotsessi muuta objektiivsemaks ja suurendaksid tulemuste usaldusväärsust.

Lisaks tuleb arvestada võimalusega, et andmetes võivad esineda ka kliiniliselt haigete patsientide analüüsitulemused, mis mõjutavad referentsväärtuste määramist. Kuigi praegune andmete eeltöötlus võib olla piisav, siis tasuks kaaluda ka rangemat andmete puhastust ning analüüsida, kas ja kuidas see mõjutab saadud tulemusi. Veel üheks potentsiaalseks probleemiks on ekstreemsete

väärtuste mõju. Väiksemate valimite puhul võivad üksikud erandid mõjutada jaotuse kuju ja seeläbi ka määratud referentsväärtuseid. Selliste mõjude hindamist võiks tulevikus analüüsida.

Samuti vajab tähelepanu tulemuste tõlgendamine. Vaja on kriitiliselt hinnata, kuidas tõlgendada rahvastiku näitel määratud referentsväärtuste võrdlust hetkel kasutusel olevate väärtustega. Kas mitte kokku langevad väärtused ei viita metoodilisele probleemile või hoopis Eesti rahvastiku eripärale? Ning vastupidi, kui väärtused langevad kokku, kas see kinnitab metoodika korrektsust või ei ole metoodika piisavalt tundlik võimalike erinevuste suhtes.

Kokkuvõttes on töö tulemused koostööpartnerite, Eesti Laborimediitsiini Ühingu spetsialistide hinnangul väärtuslikud ja kinnitavad kasutatud metoodika potentsiaali. Siiski tuleb tulemusi hinnata kriitiliselt ning metoodikat edasi arendada. Edasises uurimistöös võiks töötada välja veel objektiivsema loogika rühmade eristamiseks, täiustada andmete puhastamise kriteeriume ning arendada edasi kasutajaliidest. Samuti oleks oluline hinnata metoodika sobivust erinevate analüütide puhul. See aitaks selgitada, milliste analüütide korral on see metoodika sobilik ja milliste puhul võiks kaaluda alternatiivseid meetodeid.

## 6. Kokkuvõte

Magistritöö eesmärgiks oli määrata Eesti rahvastiku põhised referentsväärtused, pöörates tähelepanu võimalikele vanuse- ja soopõhiste erinevustele. Töö keskendus kolmele keemia ja immunoloogia analüüdile: naatrium, kreatiniin ja ferritiin. Referentsväärtuste määramine põhines kaudsel meetodil, kasutades andmeid tervise infosüsteemist ning rakendades refineR algoritmi. Lisaks töötati välja kasutajaliides referentsväärtuste määramiseks ja visualiseerimiseks. Töö eesmärk oli hinnata, kas Eesti populatsiooni põhjal saab määrata referentsväärtusi ning kuidas need erinevad seadmetootjate omadest. Lisaks uuriti, kas demograafilised tegurid, nagu sugu ja vanus, mõjutavad neid väärtusi.

Töö tulemusena leiti naatriumile, kreatiniinile ja ferritiinile referentsväärtused. Naatriumile määrati referentsvahemikuks 136–145 mmol/L. Kreatiniinile määrati soospetsiifilised väärtused. Leitud kreatiniini referentsväärtused on meestel vahemikus 59–112  $\mu\text{mol/L}$  ja naistel 47–87  $\mu\text{mol/L}$ . Ferritiini analüüdile määrati nii soo- kui ka vanusespetsiifilised väärtused. Leitud ferritiini referentsvahemikud on kuni 50-aastastel meestel 15–440  $\mu\text{g/L}$  ja naistel 3–141  $\mu\text{g/L}$ . Üle 50-aastastele meestele määrati referentsvahemikud 5–665  $\mu\text{g/L}$  ja naistele 5–395  $\mu\text{g/L}$ . Lisaks valmis töö käigus kasutajaliides, mis toetab nii andmete eeltötlust kui ka statistilist analüüsi.

Üldjoontes kattusid töö käigus määratud referentsväärtused Eestis ja rahvusvaheliselt kasutusel olevate laborite väärtustega, kuigi kohati esines mõningaid erinevusi. See viitab kasutatud meetoodika usaldusväärsusele, kuid samas toob esile ka vajaduse kriitiliselt hinnata, millal ja miks võivad erinevused tekkida. Sama meetoodikat rakendati ka uuematel andmetel lisa-analüüsis, mille tulemused olid sarnased töö põhianalüüsi tulemustega. Mõningaid erinevusi oli ferritiini referentsväärtustes, kuid üldiselt kinnitas lisa-analüüs meetoodika stabiilsust ja rakendatavust ka uuemate andmete puhul.

Püstitatud hüpoteesidele on saadud järgmised vastused:

1. Töö käigus määratud referentsväärtused kattusid enamasti olemasolevate väärtustega, kuid esines ka märgatavaid erinevusi, eriti ferritiini puhul. See viitab, et rahvastikupõhine lähenemine võib anda täpsemaid tulemusi ja sobivamaid tulemusi kohaliku rahvastiku jaoks.

2. Kreatiniini puhul oli selge erinevus sugude vahel ning ferritiini puhul mõjutasid tulemusi nii sugu kui ka vanus. Naatriumi puhul olulist erinevust ei soo ega vanuse lõikes ei ilmnenud.

Käesoleval töö on oluline praktiline tähtsus, kuna välja töötatud meetodikat on plaanis kasutada Eesti kõigi laborianalüüside referentsväärtuste määramisel ja annoteerimisel. Sellest tulenevalt mõjutab see otseselt tervishoius tehtavaid otsuseid ning patsiendile määratavat ravi. Seega on eriti oluline, et kasutatav lähenemine oleks võimalikult objektiivne ja meetoodiliselt läbimõeldud. Arutelus on põhjalikult käsitletud võimalikke ohukohti ning esitatud soovitusi nende mõju edaspidiseks vähendamiseks.

## **7. Tänuõnad**

Käesolev lõputöö on valminud uurimisprojektide PRG1844, TEM-TA72 ja TeamPerMed raames. Projekte PRG1844 ja TEM-TA72 on toetanud Eesti Teadusagentuur, Euroopa Liit läbi Euroopa Regionaalarengu Fondi (projekt nr 2021-2027.1.01.24-0444) ning Haridus- ja Teadusministeerium seoses Euroopa Horisondi meetmega "Teaming for Excellence". Soovin tänada Eesti Laborimediitsiini Ühingu spetsialiste uuritava teema selgitamisel ja väärtuslike nõuannete eest. Samuti avaldan tänu oma juhendajatele nende toetuse ja panuse eest, mis aitasid kaasa töö valmimisele.

## Viited

- Ammer T., Schützenmeister A., Prokosch H.-U., Rauh M., Rank C. M. ja Zierk J. (2021). refineR: A Novel Algorithm for Reference Interval Estimation from Real-World Data. *Scientific Reports* 11.1. Publisher: Nature Publishing Group, lk 16023. DOI: [10.1038/s41598-021-95301-2](https://doi.org/10.1038/s41598-021-95301-2).
- Bohn M. K., Bailey D., Balion C., Cembrowski G., Collier C., De Guire V., Higgins V., Jung B., Ali Z. M., Seccombe D., Taher J., Tsui A. K. Y., Venner A. ja Adeli K. (2023). Reference Interval Harmonization: Harnessing the Power of Big Data Analytics to Derive Common Reference Intervals across Populations and Testing Platforms. *Clinical Chemistry* 69.9, lk 991–1008. DOI: [10.1093/clinchem/hvad099](https://doi.org/10.1093/clinchem/hvad099).
- Box G. E. P. ja Cox D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)* 26.2, lk 211–243. DOI: [10.1111/j.2517-6161.1964.tb00553.x](https://doi.org/10.1111/j.2517-6161.1964.tb00553.x).
- Cerioti F. ja Henny J. (2008). “Are my Laboratory Results Normal?” Considerations to be Made Concerning Reference Intervals and Decision Limits. *EJIFCC* 19.2, lk 106–114.
- Clinical and Laboratory Standards Institute (2010). Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline—Third Edition. CLSI document EP28-A3c. Clinical ja Laboratory Standards Institute.
- Forsman R. W. (1996). Why is the laboratory an afterthought for managed care organizations? *Clinical Chemistry* 42.5, lk 813–816. DOI: [10.1093/clinchem/42.5.813](https://doi.org/10.1093/clinchem/42.5.813).
- Geffré A., Friedrichs K., Harr K., Concordet D., Trumel C. ja Braun J.-P. (2009). Reference values: a review. *Veterinary Clinical Pathology* 38.3, lk 288–298. DOI: [10.1111/j.1939-165X.2009.00179.x](https://doi.org/10.1111/j.1939-165X.2009.00179.x).
- Haeckel R., Adeli K., Jones G., Sikaris K. ja Wosniok W. (2023). Definitions and major prerequisites of direct and indirect approaches for estimating reference limits. *Clinical Chemistry and Laboratory Medicine (CCLM)* 61.3. Publisher: De Gruyter, lk 402–406. DOI: [10.1515/cclm-2022-1061](https://doi.org/10.1515/cclm-2022-1061).
- Ichihara K., Yamashita T., Kataoka H. ja Sato S. (2023). Critical appraisal of two Box-Cox formulae for their utility in determining reference intervals by realistic simulation and extensive real-world data analyses. *Computer Methods and Programs in Biomedicine* 242, lk 107820. DOI: [10.1016/j.cmpb.2023.107820](https://doi.org/10.1016/j.cmpb.2023.107820).
- Jian J., Pelle E. ja Huang X. (2009). Iron and Menopause: Does Increased Iron Affect the Health of Postmenopausal Women? *Antioxidants & Redox Signaling* 11.12, lk 2939–2943. DOI: [10.1089/ars.2009.2576](https://doi.org/10.1089/ars.2009.2576).

- Jones G. R. D., Haeckel R., Loh T. P., Sikaris K., Streichert T., Katayev A., Barth J. H. ja Ozarda Y. (2019). Indirect methods for reference interval determination – review and recommendations. *Clinical Chemistry and Laboratory Medicine (CCLM)* 57.1. On behalf of the IFCC Committee on Reference Intervals and Decision Limits, lk 20–29. DOI: [10.1515/cclm-2018-0073](https://doi.org/10.1515/cclm-2018-0073).
- Kallion K. (2011). Naatrium (S,P-Na). <https://www.kliinikum.ee/yhendlabor/raamat/Mn/Naatrium.pdf> (09.03.2025).
- Kim H.-Y. (2014). Statistical notes for clinical researchers: Nonparametric statistical methods: 1. Nonparametric methods for comparing two groups. *Restorative Dentistry & Endodontics* 39.3, lk 235–239. DOI: [10.5395/rde.2014.39.3.235](https://doi.org/10.5395/rde.2014.39.3.235).
- Kurstjens S., Dam A. D. van, Oortwijn E., Elzen W. P. J. den, Candido F., Kusters R., Schipper A., Kortmann Y. F. C., Herings R. M. C., Kok M., Krabbe J., Boer B. A. de, Jong A.-M. de ja Frasa M. A. M. (2025). Inconsistency in ferritin reference intervals across laboratories: a major concern for clinical decision making. *Clinical Chemistry and Laboratory Medicine* 63.3, lk 600–610. DOI: [10.1515/cclm-2024-0826](https://doi.org/10.1515/cclm-2024-0826).
- Medical Council of Canada (2023). List of normal lab values. Clinical laboratory tests and adult normal values. <https://mcc.ca/examinations-assessments/resources-to-help-with-exam-prep/normal-lab-values/> (27.03.2025).
- Meyer A., Müller R., Hoffmann M., Skadberg Ø., Ladang A., Dieplinger B., Huf W., Stankovic S., Kapoula G. ja Orth M. (2023). Comparison of three indirect methods for verification and validation of reference intervals at eight medical laboratories: a European multicenter study. *Journal of Laboratory Medicine* 47.4. Publisher: De Gruyter, lk 155–163. DOI: [10.1515/labmed-2023-0042](https://doi.org/10.1515/labmed-2023-0042).
- Reimand K. (2024). Kreatiniin seerumis/plasmas (S,P-Crea) Kreatiniin uriinis (U-Crea, dU-Crea) Hinnanguline glomerulaarfiltratsiooni kiirus (eGFR (Crea, CKD-EPI) ja eGFR (Crea, CysC, CKD-EPI)) Kreatiniinikliirens (dU-S-Crea). [https://www.kliinikum.ee/yhendlabor/wp-content/uploads/2024/08/Kreatiniin\\_v3.pdf](https://www.kliinikum.ee/yhendlabor/wp-content/uploads/2024/08/Kreatiniin_v3.pdf) (09.03.2025).
- SYNLAB Eesti (2024). Referentsväärtuste tabel. SYNLAB Eesti. <https://synlab.ee/arstile/laboriteatmik/referentsvaartused/> (30.03.2025).
- Tartu Ülikooli Kliinikumi Ühendlabor (2025). TULKHP-9.2.3v30. Laborianalüüside vastuste meditsiiniline tõlgendamine. Juhenddokument. <https://www.kliinikum.ee/yhendlabor/wp-content/uploads/2025/03/TULKHP-9.2.3v30.pdf> (13.05.2025).
- Vaagen K. (2023). Ferritiin (S,P-Fer). <https://www.kliinikum.ee/yhendlabor/ferritiin-sp-fer/> (09.03.2025).

## Lisad

### I. Referentsväärtused


Tabel 3. Põhianalüüsi käigus määratud referentsväärtused koos nende 95% usaldusintervallidega.


Analüüt	Ühik	N	Referentsväärtused		95% usaldusintervallid	
			Alumine (2,5%)	Ülemine (97,5%)	Alumisele väärtusele	Ülemisele väärtusele
Naatrium	mmol/L	46322	135,50	145,34	(135,14; 136,04)	(144,77; 145,68)
Kreatiniin mehed	$\mu\text{mol/L}$	28683	58,63	111,95	(57,05; 59,22)	(106,24; 113,14)
Kreatiniin naised	$\mu\text{mol/L}$	44101	47,32	86,53	(45,81; 49,45)	(84,08; 89,34)
Ferritiin $\leq 50$ a mehed	$\mu\text{g/L}$	2049	15,07	440,47	(8,43; 24,22)	(391,91; 511,18)
Ferritiin $\leq 50$ a naised	$\mu\text{g/L}$	6446	3,05	141,36	(2,09; 3,57)	(108,09; 157,14)
Ferritiin $\geq 51$ a mehed	$\mu\text{g/L}$	3153	5,18	664,73	(2,29; 25,10)	(572,36; 942,41)
Ferritiin $\geq 51$ a naised	$\mu\text{g/L}$	7008	4,57	394,75	(2,36; 12,02)	(297,37; 542,62)


Tabel 4. Lisa-analüüsi käigus määratud referentsväärtused koos nende 95% usaldusintervallidega.

Analüüt	Ühik	N	Referentsväärtused		95% usaldusintervallid	
			Alumine (2,5%)	Ülemine (97,5%)	Alumisele väärtusele	Ülemisele väärtusele
Naatrium	mmol/L	319433	135,76	144,10	(135,64; 135,81)	(144,05; 144,18)
Kreatiniin mehed	µmol/L	315891	59,19	112,46	(57,98; 60,94)	(107,35; 114,13)
Kreatiniin naised	µmol/L	426996	50,68	88,65	(48,73; 51,23)	(85,11; 89,34)
Ferritiin ≤ 50 a mehed	µg/L	49376	28,41	475,75	(22,89; 38,77)	(418,08; 527,64)
Ferritiin ≤ 50 a naised	µg/L	117278	4,91	168,62	(4,44; 5,38)	(165,43; 172,98)
Ferritiin ≥ 51 a mehed	µg/L	52818	16,36	682,72	(6,98; 29,42)	(567,35; 795,56)
Ferritiin ≥ 51 a naised	µg/L	106102	14,69	372,33	(6,74; 20,49)	(324,86; 433,75)

## II. Kasutajaliidese avalehe vaheleht

Referentsväärtuste analüüs 

Info 


Andmete laadimine 

**Vali CSV fail**

testandmed.csv

Päis


**Eraldaja**

Koma 


**Analüüdi nimetus**

Test


**Isiku ID veerg**

ID 


**Sugu veerg**

gender\_concept\_id 


**Vanuse veerg**

Vanus 


**Analüüsi väärtuse veerg**

Tulemus 

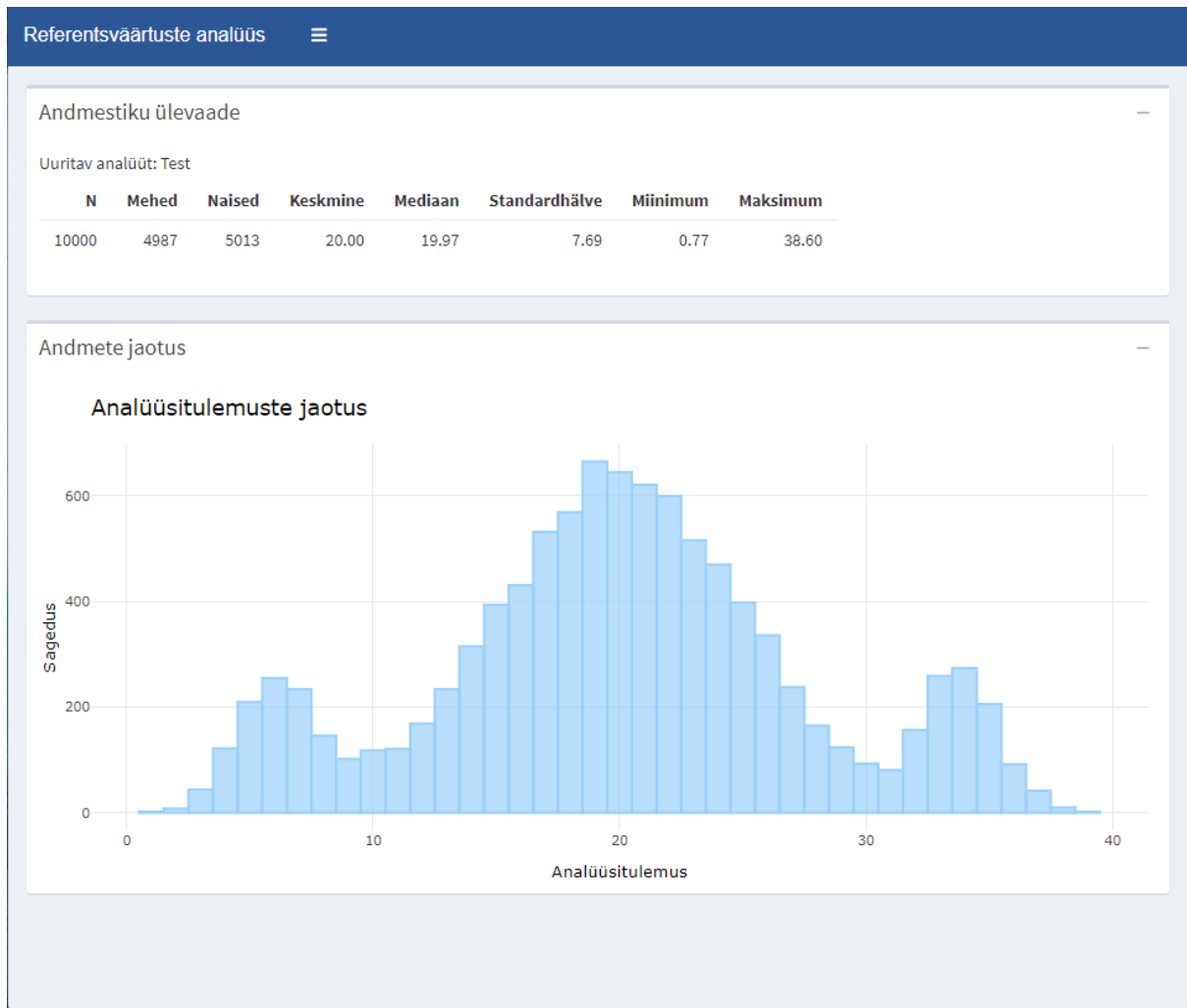
**Mehe väärtus**

8507 

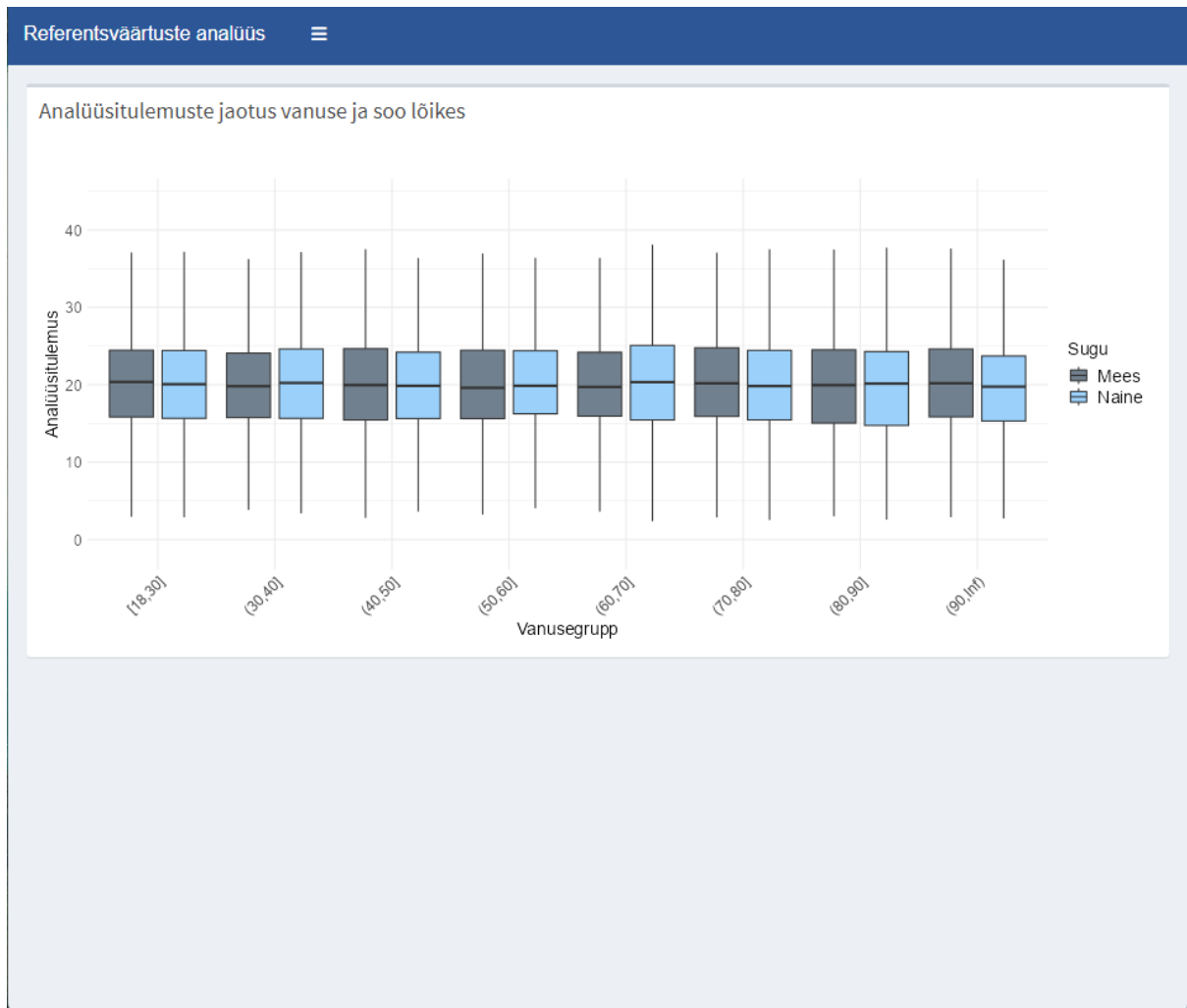
**Naise väärtus**

8532 

### III. Kasutajaliidese andmete ülevaate vaheleht



## IV. Kasutajaliidese vanuse ja soo analüüsi vaheleht



## V. Kasutajaliidese referentsväärtuste määramise vaheleht

### Referentsväärtuste analüüs

Soovitused

Uuritav analüüt: Test  
Wilcoxon'i testi p-väärtus: 0.5393  
Tulemus: Statistiliselt oluline erinevus puudub.  
Soovitus: Ei ole vaja leida soospetsifilisi referentsväärtusi.

Parameetrite valimine

Kas soo-spetsifilised referentsväärtused?

Ei  
 Jah

Millist Boxi-Coxi teisendust rakendatakse?

Üheparameetiline  
 Kaheparameetiline

Vanuse piirid

18 105

0 100 300

Arvuta referentsväärtused

Tulemused

10.4 - 29.79  
95% usaldusvahemikud: alumisele (9.91; 12.28), ülemisele (29.53; 30.51)

Test

Sagedus

Analüüsitulemus

## **VI. GitLabi projekt**

GitLabi projekti link on: [https://gitlab.cs.ut.ee/health-informatics/student-theses/paula\\_marie\\_loopere\\_lab\\_meas\\_ref\\_values](https://gitlab.cs.ut.ee/health-informatics/student-theses/paula_marie_loopere_lab_meas_ref_values)

## Litsents

### Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Paula Marie Loopere,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose  
”Laboris mõõdetavate analüütide referentsväärtuste määramine Eesti rahvastiku näitel”,  
mille juhendajad on Jaak Vilo ja Sulev Reisberg,  
reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada Tartu Ülikooli digitaalarhiivi kuni autoriõiguse kehtivuse lõppemiseni;
2. annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi kaudu Creative Commonsi 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.

Paula Marie Loopere

**15.05.2025**