

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MATEMAATIKA JA STATISTIKA INSTITUUT

Birthe Lall

**Psühhootilise häirega patsientidele mõeldud
metaboolse sündroomi riski prognoosimudeli
valideerimine Eesti geenivaramu andmetel**

Matemaatiline statistika

Bakalaureusetöö (9 EAP)

Juhendajad: PhD Maris Alver,
PhD Märt Möls

TARTU 2026

PSÜHHOOTILISE HÄIREGA PATSIENTIDELE MÕELDUD METABOOLSE SÜNDROOMI RISKI PROGNOOSIMUDELI VALIDEERIMINE EESTI GEENIVARAMU ANDMETEL

Bakalaureusetöö

Birthe Lall

Lühikokkuvõte

Käesoleva töö eesmärk oli testida Tartu Ülikooli Eesti geenivaramu andmete peal Ühendkuningriigis välja töötatud mudelit, millega on võimalik prognoosida psühhootilise häire diagnoosiga indiviididel metaboolse sündroomi kujunemise riski järgmise viie aasta jooksul. Samuti oli üheks eesmärgiks hinnata, kas psühhootilise diagnoosiga indiviidide jaoks koostatud metaboolse sündroomi riski prognoosiv mudel on sobiv kasutamiseks ka üldpopulatsioonis ehk kontrollrühma jaoks. Esmalt selgitati, mida mõeldakse psühhootiliste häirete all, mis on metaboolne sündroom ja mis mudelid on välja töötatud. Seejärel selgitati, milliste kriteeriumite alusel Eesti geenivaramu andmetel psühhootilise diagnoosiga indiviidide ja kontrollide valimid koostati. Kirjeldati ka meetodikat: mida tähendab mudeli sobivuse hindamine ning milleks kasutatakse Hosmer-Lemeshow testi. Seejärel anti ülevaade andmestikest koos kirjeldava analüüsiga ning selgitati, miks käesolevas töös testitakse ainult Ühendkuningriigis välja töötatud osalist mudelit. Viimaks analüüsitakse andmeid, esitatakse analüüsitulemused ja järeldused.

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

Märksõnad: geenivaramu, metaboolne sündroom, psühhootilised häired, Hosmer-Lemeshow test.

VALIDATION OF A METABOLIC SYNDROME RISK PREDICTION MODEL FOR PATIENTS WITH PSYCHOTIC DISORDERS USING THE ESTONIAN BIOBANK DATA

Bachelor thesis

Birthe Lall

Abstract

The aim of this study was to test a model that predicts the risk of metabolic syndrome within the next five years among individuals with a psychotic disorder. The model was developed in the UK specifically for individuals with a psychotic disorder and in this thesis tested using data from the Estonian Biobank. Additionally, we set out to investigate whether the proposed model predicts the risk of developing metabolic syndrome also in the general population (referred here as the control group). First, an overview of psychotic disorders and the metabolic syndrome was outlined, and the UK-developed prediction model was introduced. Next, the criteria used to define the sample of individuals with psychotic disorders and the control set in the Estonian Biobank were given, and the methodological background was presented, including the concept of the goodness of fit for the model and the purpose of the Hosmer-Lemeshow test. Finally, an overview of the datasets with descriptive analyses were outlined together with the rationale for testing only the partial UK-developed model. Lastly, the analysis process and results with conclusions were presented.

CERCS research specialisation: P160 Statistics, operation research, programming, actuarial mathematics.

Key Words: biobank, metabolic syndrome, psychotic disorders, Hosmer-Lemeshow test.

Sisukord

Sissejuhatus	4
1 Psühhootilised häired ja metaboolne sündroom	6
2 Metoodika	11
2.1 Mudeli sobivuse hindamine	11
2.1.1 Hosmer-Lemeshow test	12
3 Andmestike tutvustus	19
3.1 Juhtude andmestik	19
3.1.1 Kirjeldav analüüs	25
3.2 Kontrollide andmestik	29
3.2.1 Kirjeldav analüüs	29
4 Analüüsitulemused	31
4.1 Mudeli testimine juhtudel	31
4.2 Mudeli testimine kontrollrühmas	35
4.3 Mudeli hindamine juhtudel	38
4.4 Rek calibreeritud mudeli ja juhtudel hinnatud mudeli võrdlus . .	41
Kokkuvõte	42
Kasutatud allikad	44

Sissejuhatus

Metaboolne sündroom esineb psühhootilise häirega (peamiselt skisofreenia-spektrumi häirega) inimestel 30% sagedamini kui üldpopulatsioonis ning selle puudulik käsitus on omakorda seotud suurenenud suremuse riskiga nimetatud indiviidide rühmas (Alver *et al.*, 2024). Metaboolne sündroom kujutab endast kardiovaskulaarsete riskitegurite kogumit, mille hulka kuuluvad kõhuõõne rasvumine, düslipideemia, kõrge veresuhkur ja kõrgvererõhutõbi (Ma *et al.*, 2009). Seetõttu on oluline ja vajalik täpsustada ning edendada metaboolse sündroomi ennustus- ja ennetusstrateegiaid (Alver *et al.*, 2024). On näidatud, et üldpopulatsioonile mõeldud metaboolse sündroomi riskihindamise mudelid ei ole psühhootilise häire diagnoosiga inimestele sobilikud, kuna ei hinda riski piisavalt täpselt. Hiljuti arendati Ühendkuningriigis nimetatud patsientide rühmale metaboolse sündroomi riski mudel, kuid seda ei ole Eestis veel valideeritud ega kasutusele võetud. (Perry *et al.*, 2022)

Bakalaureusetöö eesmärk on testida ja vajadusel modifitseerida Eesti andmete peal nimetatud mudelit, millega on võimalik prognoosida psühhootilise häire diagnoosiga indiviididel metaboolse sündroomi kujunemise riski. Algse mudeli autorite valideerimisanalüüsis oli keskmine jälgimisaeg 5,18 aastat (ALSPAC valim) (Perry *et al.*, 2021). Seetõttu soovitakse käesolevas töös kontrollida mudelit, mis prognoosib psühhootilise diagnoosiga indiviididel metaboolse sündroomi tekkeriski järgmise viie aasta jooksul. Samuti on üheks eesmärgiks hinnata, kas psühhootilise diagnoosiga indiviidide jaoks koostatud metaboolse sündroomi riski prognoosiv mudel on sobiv kasutamiseks ka üldpopulatsiooni kuuluvate indiviidide ehk kontrollide jaoks. Andmed pärinevad Tartu Ülikooli Eesti geenivaramust.

Töö esimene peatükk on psühhootilistest häiretest ja metaboolsest sündroomi

mist. Teine peatükk on metoodikast: mudeli sobivuse hindamisest, täpsemalt Hosmer-Lemeshow testist. Kolmandas peatükis antakse ülevaade andmestikest koos kirjeldava analüüsiga. Viimases peatükis analüüsitakse Eesti geonivaramu andmeid, esitatakse analüüsitulemused ja järeldused.

Käesolevas töös kasutati tarkvara R koodi parandamiseks, ingliskeelse teksti tõlkimiseks ja abivahendina teksti keeleliseks viimistlemiseks tehisaru ChatGPT abi (OpenAI, [2026](#)).

1 Psühhootilised häired ja metaboolne sündroom

Käesolevas töös käsitletakse psühhootiliste häirete all skisofreeniaspektri häiret ning bipolaarset sündroomi. Skisofreeniale on iseloomulik lai sümptomispekter, sealhulgas positiivsed sümptomid (nt hallutsinatsioonid, organiseerimatu kõne, luulumõtted), negatiivsed sümptomid (motivatsioonipuudus ning kõne- ja emotsionaalse väljenduse vähenemine) ja kognitiivsed sümptomid (tähelepanu, pikaajalise mälu ning täidesaatvate funktsioonide puudujäägid) (Jauhar, Johnstone ja McKenna, 2022; Lim *et al.*, 2026). Psühhootiliste häirete kujunemine on seotud nii geneetiliste kui ka keskkondlike riskiteguritega (Alver *et al.*, 2024) ning sümptomite avaldumine algab tavaliselt noores täiskasvanueas ehk 14–30-aastaselt, algusvanuse mediaan on 25 aastat (Lim *et al.*, 2026). Haiguse kulg ja tagajärjed on märkimisväärsed: skisofreenia diagnoosiga inimeste eluiga on keskmiselt 15 aastat lühem kui üldpopulatsioonis (Alver *et al.*, 2024).

Bipolaarne häire on raske krooniline meeleluuhäire, mida iseloomustavad maania ja hüpomaania episoodid ning vahelduvad või põimuvad depressiooniepisoodid. Maniakaalsed ja hüpomaniakaalsed episoodid on ajutised seisundid, mida iseloomustavad kõrgenenud meeleolu ning motoorse aktiivsuse tõus ja mis erinevad oma raskusastme ning kestuse poolest. (Grande *et al.*, 2016) Uuringute kohaselt on bipolaarse häire diagnoosiga indiviididel, sarnaselt skisofreeniaspektri häire diagnoosiga inimestele, lühem eluiga, ligikaudu 10–20 eluaasta võrra (McIntyre *et al.*, 2020).

Skisofreenia ja teiste psühhootiliste häirete ravis kirjutatakse patsientidele tihti välja antipsühhootilisi ravimeid, kuid nende toime ei ole alati piisav ning nendega võivad kaasneda olulised kõrvaltoimed, näiteks kehakaalu tõus

ja suurenenud risk metaboolse sündroomi kujunemiseks (Lim *et al.*, 2026).

Nagu eelnevalt mainitud, kujutab metaboolne sündroom endast kardiovaskulaarsete riskitegurite kogumit, mille hulka kuuluvad kõhuõõne rasvumine, düslipideemia, kõrge veresuhkur ja kõrgvererõhutõbi (Ma *et al.*, 2009). Järgmises tabelis (vt tabel 1) on esitatud metaboolse sündroomi diagnoosikriteeriumid koos piirväärtustega, mis on võetud vastavalt Perry *et al.* (2022) artiklile ning mida kasutatakse ka käesoleva töö analüüsis.

Tunnus	Kriteerium
Kehamassiindeks	> 29,9
Triglütseriidid	> 1,7 mmol/l
Veresuhkur	> 5,6 mmol/l
HDL-kolesterool	< 1,03 mmol/l (mehed), < 1,29 mmol/l (naised)
Süstoolne vererõhk	> 130 mmHg

Tabel 1: Metaboolse sündroomi kriteeriumid.

Metaboolne sündroom diagnoositakse juhul, kui indiviidil ületab vähemalt kolm tunnust tabelis 1 esitatud piirväärtuse. (Perry *et al.*, 2022) Metaboolne sündroom kujutab endast olulist terviseriski, olles seotud nii suurenenud haigestumuse kui ka suremusega, sealhulgas suurema II tüüpi diabeedi ja kardiovaskulaarsete haiguste riskiga (Ma *et al.*, 2009).

Metaboolse sündroomi esinemissagedus on psühhootilise häirega inimestel 30% kõrgem kui üldpopulatsioonis ning selle ebapiisav käsitlemine on omakorda seotud suurenenud suremuse riskiga. Seetõttu on oluline ja vajalik arendada tõhusamaid ennustus- ja ennetusmeetodeid. (Alver *et al.*, 2024)

Ühendkuningriigi andmeid kasutades on välja töötatud kaks mudelit, millega on võimalik psühhootiliste häiretega noortel inimestel (vanuses 16–35) hinnata metaboolse sündroomi kujunemise riski (Perry *et al.*, 2022). Mudelid hinnati 651-st vaatlusest koosneva andmestiku peal ning seejärel valideeriti

eraldi andmestikul, mis koosnes 510-st vaatlusest. Lisaks viidi läbi tundlikkusanalüüs, kasutades 505 vaatlusega valimit. Mõlema mudeli korral oli AUC suurem kui 0,70, mis näitab üsna head eristusvõimet. (Perry *et al.*, 2021) Mudelid ja nende parameetrid on nähtavad järgmises tabelis (vt tabel 2).

Tunnus	Täismudel	Osaline mudel
Vabaliige	-6,439813	-6,973829
Vanus aastates (pidev tunnus)	0,006233226	0,00633115
Mustanahaline/Aafrika-Kariibi päritolu (jah/ei)	0,004258861	0,07548129
Aasia/muu päritolu (jah/ei)	0,211217746	0,29285950
Sugu mees (jah/ei)	0,222300765	0,31460036
KMI (pidev tunnus)	0,141186241	0,16912161
Suitsetamise staatus (suitsetaja/mitte-suitsetaja)	0,153691193	0,24751854
Määratud metaboolselt aktiivne antipsühhootikum (jah/ei)	0,497552758	0,60013558
HDL-kolesterool (pidev tunnus)	-0,399013329	—
Triglütseriidid (pidev tunnus)	0,343528440	—

Tabel 2: Mudelite kordajad.

Tunnused valiti kliiniliste teadmiste põhjal, arvestades nende tõenäolist kliinilist kasulikkust ning mõõtmiste koormust ja vastuvõetavust patsiendi seisukohalt. Osaline mudel arendati juhaks, kui biokeemilised tulemused ei pruugi olla saadaval. (Perry *et al.*, 2022)

Selleks, et liikuda edasi rahvusvaheliselt rakendatava metaboolse sündroomi kujunemise riski prognoosiva mudeli suunas, on mudelit valideeritud ka kahes sõltumatus Euroopa valimis. Kasutati Šveitsi ja Hispaania kohortide andmeid, kus kaasati osalejaid vanuses 16–35, kellel ei esinenud uuringu alguses metaboolset sündroomi ning kelle jälgimisperiood oli 1–6 aastat. Tuvastati, et võrreldes osalise mudeliga, eristas täismudel metaboolse sündroomi esinemist haiguse mitteesinemisest paremini. Näiteks esineb esmaste psüh-

hoosiepisoodiga ravita patsientide seas sageli metaboolseid kõrvalekaldeid, sealhulgas insuliiniresistentsust ja düslipideemiat, seda isegi normaalse kehamassiindeksi korral. Kuna kõrgeenenud triglütseriidide ja HDL-kolesterooli suhe viitab insuliiniresistentsusele ning täielik mudel võimaldab neid kõrvalekaldeid tuvastada, samas kui osaline mudel seda ei võimalda, võib see selgitada täismudeli paremat eristusvõimet. (Perry *et al.*, 2022)

Käesoleva töö andmed pärinevad Tartu Ülikooli Eesti geenivaramust. Vaatluse alla võeti psühhootilise diagnoosiga inividid, s.t inimesed, kellel on skisofreeniaspektri häire (üks RHK-10 F20-F29 kirjetest), bipolaarne sündroom (F31) ja/või maania (F30). Vastavalt geenivaramu andmekoosseisule 2025v3 koosneb uuritav valim 1633 indiviidist, kes vastavad järgmistele kriteeriumitele:

- olemas vähemalt üks RHK-10 F20-F31 kirjetest,
- vähemalt üks diagnoos on määratud psühhiaatri poolt,
- esimese diagnoosi või antipsühhootilise/meeleolu stabilisaatorite ravi määramise vanus on vahemikus 14–50,
- vähemalt üks ambulatoorne või statsionaarne diagnoos on registreeritud Eesti Tervisekassa andmetes,
- esimene diagnoos ei tohi olla varasem kui 2005.

Et hinnata metaboolse sündroomi kujunemise riski prognoosiva mudeli enustusvõimet üldpopulatsioonis, võeti iga psühhootilise häire diagnoosiga indiviidile kuni kolm kontrollindiviidi vastavalt soole ja sünniaastale ning järgmistele kriteeriumitele:

- puudus psühhootilise häire diagnoos ning muud psüühika- ja käitumishäirete RHK-10 kirjed (RHK-10 F kirje),
- olemas andmed suitsetamise ning kehamassiindeksi (edaspidi KMI) kohta,
- elus ja vaadeldav sama perioodi jooksul kui juhtum (või kuni surmani, kui see saabus selle aja jooksul),
- puudusid metaboolse sündroomi tunnused enne sobitatud juhtumi haiguse avaldumist (esimese diagnoosi või antipsühhootilise/meeleolu stabilisaatorite ravi määramise hetk),
- oli olemas vähemalt üks KMI mõõtmine kuni 3 aastat enne sobitatud juhtumi haiguse avaldumist.

Uuring teostati väljastuse U60 raames. Projekti läbiviimiseks on Eesti bioetika ja inimuuringute nõukogult saadud kooskõlastus 26. aprillil 2025 (otsus nr 1.1-12/1154). Jätkutaotlusele anti kooskõlastus 9. oktoobril 10.2025 (otsus nr 1.1-12/2537).

2 Metoodika

Käesolevas peatükis kirjeldatakse kasutatud metoodikat.

2.1 Mudeli sobivuse hindamine

Alapeatükk on kirjutatud raamatu „Applied Logistic Regression“ (Hosmer, Lemeshow ja Sturdivant, 2013) põhjal.

Olgu uuritav tunnus y binaarne, kus $y_i = 1$ tähendab, et i -ndal vaatlusel sündmus toimus, ja $y_i = 0$, et sündmus ei toimunud. Olgu meil sobitatud mudel, mis sisaldab a sõltumatut muutujat. Kasutades sobitatud mudeli logiti hinnangut

$$\hat{g}(\mathbf{x}_i) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_a x_{ia},$$

kus $\mathbf{x}'_i = (x_{i1}, x_{i2}, \dots, x_{ia})$ on i -nda vaatluse seletavate tunnuste vektor, saab leida hinnatud tõenäosuse kujul

$$\hat{p}_i = \frac{e^{\hat{g}(\mathbf{x}_i)}}{1 + e^{\hat{g}(\mathbf{x}_i)}}.$$

Logistilise regressioonimudeli hindamisel kontrollitakse esmalt mudeli üldist statistilist olulisust ning mudelisse kaasatud tunnuste olulisust ja funktsionaalset vormi. Mudeli sobivuse hindamisel vaadatakse, kuivõrd mudeli prognoositud tõenäosused vastavad tegelikele andmetele. Mudel sobib, kui tegelikud ja prognoositud väärtused on keskmiselt samad ning erinevused nende vahel ei moodusta süstemaatilist mustrit, vaid on väikesed ja mudeli eeldustega kooskõlas.

Mudeli sobivuse hindamine on absoluutne, mitte suhteline, kuna prognoositud väärtusi võrreldakse otseselt vaadeldud väärtustega. Vaadeldud väärtusi võib tõlgendada kui ideaalse mudeli väljundit, millele hinnatav mudel püüab võimalikult lähedale jõuda. Mudeli sobivust saab hinnata mitmel eri viisil, sealhulgas kasutades Hosmer–Lemeshow testi, nagu ka antud töös tehti. Nimetatud testi käsitletakse järgnevas alapeatükis detailsemalt.

2.1.1 Hosmer-Lemeshow test

Hosmer-Lemeshow testi kasutatakse selleks, et vaadata, kui hästi vastab logistilise regressiooni mudeli prognoositud tõenäosus tegelikele tulemustele ehk kui hästi mudel sobib andmetega. Testi hüpoteesid on:

$$H_0 : \text{mudel sobib andmetega}$$

$$H_1 : \text{mudel ei sobi andmetega}$$

Hosmer-Lemeshow statistiku leidmiseks ja testi tegemiseks pakuti välja gruppimine mudeliga hinnatud tõenäosuste põhjal. Selleks eeldatakse, et iga vaatluse kohta on olemas mudeliga hinnatud tõenäosus, mis järjestatakse kasvavalt ja mida saab seejärel jaotada gruppidesse. Pakuti välja kaks gruppimisviisi:

1. Hinnatud tõenäosuste protsentiilide alusel - iga grupp sisaldaks umbes sama palju vaatlusi.
2. Hinnatud tõenäosuse fikseeritud väärtuste alusel - grupid kindlate vahemikega, aga igas grupis võib olla erinev arv andmeid.

Kui kasutada $g = 10$ gruppi ning n on kõigi vaatluste arv kokku, siis esimese meetodiga sisaldaks iga grupp ligikaudu $n/10$ vaatlust. Esimeses grupis oleks

kõige väiksema hinnatud tõenäosusega vaatlused ning viimases kümnendas kõige suurema hinnatud tõenäosusega vaatlused. Kui kasutada $g = 10$ gruppi, siis teise grupeerimisviisi korral määratakse jaotuspiirid väärtuste $k/10$ järgi ($k = 1, 2, \dots, 9$) ning iga grupp sisaldab kõiki vaatlusi, mille hinnatud tõenäosused jäävad kahe järjestikuse jaotuspiiri vahele. Näiteks koosneks esimene grupp kõikidest vaatlustest, mille prognoositud tõenäosused oleksid $\leq 0,1$, ning kümnes grupp koosneks kõigist vaatlustest, mille hinnatud tõenäosused on $> 0,9$.

Tähistagu c_k k -ndasse gruppi kuuluvate vaatluste indeksite hulka ning olgu $n_k = |c_k|$ k -ndasse gruppi kuuluvate vaatluste arv. Iga grupi kohta leitakse vaadeldud ja oodatud sagedused. Vaadeldud toimunud sündmuste arv k -ndas grupis on

$$o_{1k} = \sum_{i \in c_k} y_i$$

ning vaadeldud mittetoimunud sündmuste arv on

$$o_{0k} = n_k - o_{1k}.$$

Oodatud toimunud sagedus leitakse summeerides kõikide gruppi kuuluvate vaatluste mudeliga hinnatud tõenäosused:

$$\hat{e}_{1k} = \sum_{i \in c_k} \hat{p}_i.$$

Oodatud mittetoimunud sagedus saadakse liites kõigi gruppi kuuluvate vaat-

luste i korral üks miinus hinnatud tõenäosus $(1 - \hat{p}_i)$ ehk

$$\hat{e}_{0k} = \sum_{i \in c_k} (1 - \hat{p}_i) = n_k - \hat{e}_{1k}.$$

Mõlema eelnevalt mainitud grupeerimise korral saadakse Hosmer-Lemeshow statistik $\hat{C}$ arvutades Pearsoni hii-ruut statistik $g \times 2$ tabeli põhjal, kus on esitatud vaadeldud ja mudeli põhjal oodatud sagedused. Väikese p-väärtuse korral lükatakse nullhüpotees tagasi ning järeldatakse, et mudel ei sobi andmetega piisavalt hästi. Statistiku $\hat{C}$ valem on järgmine:

$$\hat{C} = \sum_{k=1}^g \left[\frac{(o_{1k} - \hat{e}_{1k})^2}{\hat{e}_{1k}} + \frac{(o_{0k} - \hat{e}_{0k})^2}{\hat{e}_{0k}} \right]. \quad (1)$$

Valem 1 asemel võib kasutada ka sellist valemit:

$$\hat{C} = \sum_{k=1}^g \frac{(o_{1k} - n_k \bar{p}_k)^2}{n_k \bar{p}_k (1 - \bar{p}_k)}, \quad (2)$$

mis on saadud algebralise teisendamisega ning kus $\bar{p}_k$ on k -ndas grupis olevate vaatluste hinnatud tõenäosuste keskmine, st

$$\bar{p}_k = \frac{1}{n_k} \sum_{i \in c_k} \hat{p}_i.$$

Täiendavad uuringud Hosmer *et al.* poolt on näidanud, et esimest grupeerimise meetodit tuleks eelistada teisele, sest siis jälgib teststatistiku jaotus nullhüpoteesi kehtides paremini χ^2 -jaotust vabadusastmetega $g - 2$. Seega, kui ei ole märgitud teisiti, siis Hosmer-Lemeshow statistik on saadud protsentilide alusel grupeerides.

Kui valimi suurus on alla 1000, soovitatakse kasutada kümmet gruppi. An-

tud töös on Hosmer-Lemeshow testi kasutamisel kasutatud samuti kümme gruppi, $g = 10$.

Kui mõnes lahtris on oodatud sagedused väikesed (< 5), võib tabeli mõningaid järjestikuseid ridu ühendada, et neid suurendada ja vähendada vabadusastmete arvu. Samas ei pea seda alati tegema, kuna ka väiksemate sageduste korral võivad tulemused olla piisavalt usaldusväärsed.

Järgnevalt on mõned märkused statistiku $\hat{C}$ arvutamise kohta. Kui mudel prognoosib sama tõenäosuse mitmele vaatlusele, siis võib tekkida olukord, kus sama hinnatud tõenäosusega vaatlused satuvad erinevatesse gruppidesse. Sellisel juhul sõltub statistiku $\hat{C}$ väärtus teatud määral sellest, kuidas need võrdsed väärtused detšiilidesse jaotatakse.

Erinevad tarkvaralised lahendused võivad võrdseid hinnatud tõenäosusi detšiilidesse jaotada erineval viisil. See ei tohiks üldjuhul probleeme tekitada, kui just erinevaid hinnatud tõenäosuse väärtusi on nii vähe, et kõigi võrdsete väärtuste ühte gruppi paigutamine põhjustab gruppide suurustes suure tasakaalutuse või tekitab olukorra, kus grupe on märgatavalt vähem kui 10. Sellisel juhul võib arvutatud $\hat{C}$ statistik olla eri pakettide korral üsna erinev. Lisaks, kui $\hat{C}$ arvutamiseks on kasutatud liiga vähe grupe, on oht, et test ei ole piisavalt tundlik, et eristada vaadeldud sagedusi oodatavatest. Hosmer *et al.* väidab, et kui jagada vaatlused vähem kui kuude gruppi, siis leiab Hosmer-Lemeshow test peaaegu alati, et mudel sobib andmetega.

Hosmer-Lemeshow statistiku puuduseks on veel see, et grupeerimise käigus võidakse väikese arvu üksikvaatluste tõttu jätta märkamata olulise kõrvalekalde mudeli sobivusest. Seetõttu soovitavad Hosmer *et al.* enne mudeli lõplikku sobivaks tunnistamist analüüsida üksikvaatluste jääke ja asjakohaseid diagnostilisi näitajaid.

Järgnevalt on toodud näide Hosmer-Lemeshow statistiku arvutamise kohta.

Taheti katsetada, kui hästi toimib uus ravim inimeste peal. Katsesse kaasati kokku 98 inimest, kellele manustati erinevates doosides ravimit. Erinevaid doose oli 15. Mõne aja möödudes hinnati, kui paljudel patsientidel ravi toimis (ehk sündmus $y = 1$). Seejärel koostati logistiline regressioonimudel, kus seletavaks tunnuseks oli doos ning mille abil hinnati iga vaatluse korral tõenäosust, et ravi toimib:

$$\text{logit}(p) = -0,8181 + 0,0036 \cdot \text{doos}.$$

Tulemused on esitatud tabelis 3.

Grupp	Doos	Kokku	Paljudele mõjus	Mõju osakaal	Mudeli prognoos
1	30	7	3	0,43	0,330
2	60	9	2	0,22	0,354
3	90	2	2	1,00	0,380
4	120	5	2	0,40	0,406
5	150	10	4	0,40	0,433
6	180	7	1	0,14	0,460
7	210	3	2	0,67	0,487
8	240	11	6	0,54	0,514
9	270	2	0	0,00	0,541
10	300	1	1	1,00	0,568
11	330	6	6	1,00	0,595
12	360	12	10	0,83	0,621
13	390	10	4	0,40	0,646
14	420	5	3	0,60	0,671
15	450	8	5	0,63	0,695

Tabel 3: Näidisülesande andmed.

Testime kasutades Hosmer-Lemeshow testi, kas mudel sobib andmetega. Selleks järjestame prognoositud tõenäosused kasvavalt ning jaotame need kümnesse ligikaudu võrdse suurusega gruppi (detsiilidesse). Seejärel koostame

tabeli (vt tabel 4), kus iga grupi kohta on esitatud sinna kuuluvate vaatluste hinnatud tõenäosuste vähim ja suurim väärtus. Kuna mitmel vaatlusel on sama prognoositud tõenäosus, siis võib sama tõenäosus olla mitmes grupis. Iga grupi kohta on samuti esitatud vaadeldud ja oodatud sagedused. Vaadeldud sagedused on saadud liites kokku vastavasse tõenäosusvahemikku kuuluvate vaatluste tegelikud tulemused (0/1 väärtused), kus 1 tähistab sündmuse esinemist ehk ravi toimimist. Näiteks esimesse gruppi kuulub 10 kõige väiksema hinnatud tõenäosusega vaatlust, mille prognoositud tõenäosused on vahemikus 0,330–0,354. Neist 5 puhul ravi toimis ja 5 puhul mitte, seega vaadeldud väärtused on vastavalt 5 ja 5. Oodatud sagedus on arvutatud summeerides vastavasse tõenäosusvahemikku kuuluvate vaatluste mudeli poolt hinnatud tõenäosused. Oodatud sagedus (ravi toimis) on saadud nii:

$$\hat{e}_{1,1} = 7 \cdot 0,33 + 3 \cdot 0,354 = 3,37,$$

kus esimene indeks 1 tähistab sündmust ning teine indeks esimest hinnatud tõenäosuste põhjal moodustatud gruppi. Oodatud sagedus, kus ravi ei toiminud, on saadud nii:

$$\hat{e}_{0,1} = 10 - 3,37 = 6,63.$$

$\hat{p}$ vahemik	Vaadeldud		Oodatud		Kokku
	o_{0k}	o_{1k}	$\hat{e}_{0k}$	$\hat{e}_{1k}$	
0,330–0,354	5	5	6,63	3,37	10
0,354–0,406	6	4	6,30	3,70	10
0,406–0,433	6	4	5,75	4,25	10
0,433–0,460	9	1	5,49	4,51	10
0,460–0,514	2	8	4,94	5,06	10
0,514–0,600	6	4	4,51	5,49	10
0,600–0,621	0	10	3,87	6,13	10
0,621–0,646	3	7	3,66	6,34	10
0,646–0,671	6	3	3,08	5,92	9
0,671–0,695	4	5	2,77	6,23	9

Tabel 4: Näidisülesande vaadeldud ja oodatud sagedused.

Nüüd saame arvutada Hosmer-Lemeshow statistiku tabeli 4 andmete ning valemi 1 põhjal:

$$\hat{C} = \sum_{k=1}^{10} \left[\frac{(o_{1k} - \hat{e}_{1k})^2}{\hat{e}_{1k}} + \frac{(o_{0k} - \hat{e}_{0k})^2}{\hat{e}_{0k}} \right].$$

Saame, et $\hat{C} = \left[\left(\frac{(5-3,37)^2}{3,37} + \frac{(5-6,63)^2}{6,63} \right) + \dots + \left(\frac{(5-6,23)^2}{6,23} + \frac{(4-2,77)^2}{2,77} \right) \right] = 22,08227$.

Kuna vahemikke oli 10 ehk $g = 10$, siis vabadusastmeid on $g-2 = 8$. Olulisuse nivoole 0,05 ja vabadusastmele 8 vastab hii-ruudu jaotustabelis ümardatud väärtus 15,51. Kuna $\hat{C} = 22,08 > 15,5$, siis kummutame nullhüpoteesi ehk mudel ei sobi andmetega. Seega võib järeldada, et prognoositud tõenäosused ei vasta tegelikkusele ja seos doosi ning ravi toimimise vahel erineb mudeli poolt kirjeldatavast seosest.

3 Andmestike tutvustus

Käesolevas peatükis antakse ülevaade lõputöös kasutatud andmestikest.

Mudelit hinnati kahe andmestiku, nii juhtude kui kontrollide peal.

Kontrollitav mudel prognoosib riski, et seni metaboolse sündroomi diagnoosita psühhootilise häirega indiviid saab järgnevatel aastatel metaboolse sündroomi. Mudeli valideerimisel autorite enda poolt oli keskmine jälgimisaeg 5,18 aastat (ALSPAC valim) (Perry *et al.*, 2021). Seetõttu soovitakse käesolevas töös kontrollida mudelit, mis prognoosib psühhootilise diagnoosiga indiviididel metaboolse sündroomi tekkeriski järgmise viie aasta jooksul.

Geenivaramu andmed pole otse, töötlemata kujul, sellise analüüsi teostamiseks sobivad. Osade inimeste jälgimisaeg on palju pikem kui 5 aastat, osadel lühem. On inimesi, kellel mõne tunnuse mõõtmistulemus on pärit ajast peale metaboolse sündroomi diagnoosi saamist (ja neid vaatluseid ei tohiks tulevikku prognoosivas mudelis kasutada). Seega tuleks esmalt moodustada tuleviku prognoosimiseks mõeldud mudeli valideerimiseks või uuesti hindamiseks sobiv andmestik või andmestikud Eesti geenivaramu andmete põhjal.

3.1 Juhtude andmestik

Juhtude andmestiku kokkupanemine põhines kuuel eri andmestikul, mis olid ette antud:

- Psühhootilise häire diagnoosiga indiviidide andmestik, mis oli aluseks andmete ühendamisele. See sisaldas analüüsi seisukohalt olulisi tunnuseid, nagu sugu, psühhootilise haiguse avaldumise kuupäev ning vanus psühhootilise haiguse avaldumise hetkel. Edasisel töötlusel lisati sellele andmestikule järk-järgult tunnuseid juurde.

- Psühhootilise häire diagnoosiga indiviidide ja kontrollide suitsetamise andmestik.
- Psühhootilise häire diagnoosiga indiviidide ja kontrollide RHK-10 kirjete andmestik.
- Psühhootilise häire diagnoosiga indiviidide ja kontrollide KMI andmestik.
- Psühhootilise häire diagnoosiga indiviidide ja kontrollide kliinilise vere ja vererõhu näitude andmestik.
- Andmestik, mis sisaldab iga psühhootilise diagnoosiga indiviidi kohta tunnuseid, mis kirjeldavad metaboolset aktiivsete, metaboolset mitteaktiivsete ning kõigi antipsühhootiliste ravimite ja meeolelu stabilisaatorite kasutamise proportsiooni jälgimisperioodil. Kuigi Ühendkuningriigis väljatöötatud mudelis määrati ravimikasutus binaarse tunnusega, kasutasime Eesti geenivaramu andmestikus ravimikasutuse proportsiooni jälgimisaja jooksul, et kirjeldada täpsemalt indiviidi pikaajalist ravimite võtmist.

Andmestikus, mis sisaldas iga psühhootilise diagnoosiga indiviidi kohta erinevate antipsühhootiliste ravimite ja meeolelustabilisaatorite kasutamise proportsioone jälgimisperioodil, defineeriti ravimite kasutus vastavalt ravimiostu andmetele, mis on kättesaadavad Eesti geenivaramust Tervisekassa andmestikust. Igale ravimile tuletati ravil olemise aeg vastavalt ostetud pakside arvule, paki suurusele ja doosile. Ülekattuvad perioodid liideti kokku ning igale indiviidile arvutati proportsioon jagades ravil oldud päevad jälgimisperioodi päevade arvuga. Seda tehti nii kõikidele antipsühhootilistele ravimitele ja meeolelu stabilisaatoritele kui ka eraldi kaalutõusuga seotud ravimitele,

mille hulka kuuluvad järgmised ravimid: kvetiapiin, olansapiin, aripiprasool, klosapiin, risperidoon, haloperidool, amisulpriid, sertindool, flupentiksool, perfenasiin, ziprasidoon, levomepromasiin, zuklopentiksool, kloorprotikseen, sulpiriid, kariprasiin, melperoon, kloorpromasiin, flufenasiin ja paliperidoon. (Alver *et al.*, 2024)

Esimeses etapis koondati kokku kõik vajalikud andmed üle kuue andmekoogu. Seejärel tuvastati, kellel esines metaboolne sündroom enne psühhootilise haiguse avaldumist. Kuna töös vaatluse all oleva mudeli eesmärk oli hinnata nimetatud sündroomi avaldumist, tuli nimetatud inimesed lõplikult valimist eemaldada. Lõpuks hinnati koondandmestikul, kellel esines metaboolne sündroom pärast psühhootilise haiguse avaldumist. Jälgimisaeg oli viis aastat või kuni surmani.

Kõigepealt tuli psühhootilise diagnoosiga indiviidide andmestikust (kokku 1633 vaatlust) valida inimesed, kes olid peale psühhootilise haiguse avaldumist olnud jälgimise all vähemalt 5 aastat või 5 aasta jooksul surnud. Kokku oli selliseid inimesi 1388, kellest 7 surid viie aasta jooksul. Seejärel leiti KMI andmestikust igale indiviidile KMI näit, mis oli mõõdetud maksimaalselt 3 aastat enne psühhootilise haiguse avaldumist. See oli olemas 517 inimesel. Kui indiviidil oli selles ajavahemikus mitu KMI mõõtmist, eelistati normi ületavaid väärtusi ($KMI > 29,9$), valides neist kõige hilisem. Kui normi ületavaid väärtusi ei esinenud, valiti alguspunktile lähim KMI näit. 517 inimesest 88-l oli KMI näit normist suurem. Andmestikust valiti indiviidi identifikaator, KMI ning veerg, mis ütles, kas inimesel oli KMI üle normi, ning need tunnused lisati psühhootilise diagnoosiga indiviidide algandmestikule.

Pärast seda võeti vaatluse alla metaboolse sündroomi riskifaktorite andmestik, mis sisaldas kliinilise keemia ja vererõhu mõõtmistulemusi. Need on haiglates ja terviseasutustes mõõdetud kliinilise vere analüüsitulemused, mis on

liidetud Eesti geenivaramu andmekoguga. Vaatluse alla võeti triglütseriidid (edaspidi TG), veresuhkru ehk glükoositas, HDL-kolesterool (edaspidi HDL) ja süstoolne vererõhk (edaspidi SBP) (vt tabel 1). Nende näitajate puhul kasutati sama lähenemist nagu KMI mõõtmiste korral ehk kui indiviidil oli vaadeldavas ajavahemikus mitu näitu, eelistati normi ületavaid väärtusi, valides neist alguspunktile kõige lähem. Kui normi ületavaid väärtusi ei esinenud, valiti alguspunktile lähim näit. Tuletatud andmestik lisati samuti algandmestikule. Iga indiviidi kohta säilitati üks mõõtmine vastavalt eelkirjeldatud kriteeriumitele ning puuduvate väärtuste korral eeldati, et puudumine viitab mõõtmise mitteteostamisele kõrvalekalde kahtluse puudumise tõttu. Seetõttu asendati puuduvad väärtused normis olevate näitude keskmisega. Keskmise arvutamisel leiti esmalt iga psühhootilise diagnoosiga indiviidi kohta tema normivahemikku jäävate mõõtmiste keskmine ning seejärel arvutati nende keskmiste põhjal üldkeskmine.

Metaboolse sündroomi hindamiseks võeti vaatluse alla ka järgnevad haigusdiagnoosid RHK-10 kirjete põhjal:

- E11 (insuliinisõltumatu suhkurtõbi),
- E78 (lipoproteiiniainevahetuse häired ja muud lipideemiad),
- E66 (rasvumus),
- I10 (kõrgvererõhktõbi),
- I20-I25 (südame isheemiatõved),
- I61 (peaaajusisene verevalum),
- I63 (peajuinfarkt),

- I64 (täpsustamata kas hemorraagia või infarktitekkene insult ehk rabandus).

Lisaks märgiti iga psühhootilise häire diagnoosiga indiviidile, kas ta on või ei ole kunagi suitsetanud, vastavalt „suitsetanud kunagi/praegu“ või „mitte kunagi“.

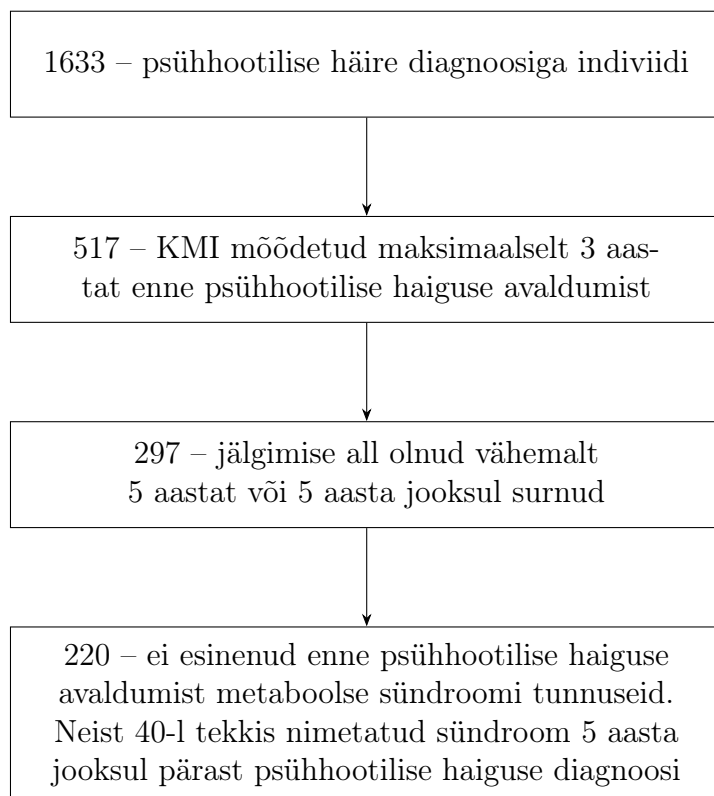
Kokkukoondatud andmestiku põhjal hinnati, kellel esines metaboolne sündroom enne psühhootilise haiguse avaldumist. Selleks lähtuti metaboolse sündroomi defineerimise kriteeriumist, mille järgi peab vähemalt kolm näitu olema normist kõrgemad. Samuti lugesime metaboolse sündroomina olukorda, kui indiviidil esines vähemalt üks ülalnimetatud RHK-10 kirje. Kui kliiniliste mõõtmiste puhul olid vaatluse all kõik näidud maksimaalselt kolm aastat enne psühhootilise haiguse avaldumist, siis RHK-10 kirjete puhul ajaraamistikku ei defineeritud. Oluline oli, et need oleksid inimesel enne psühhootilise haiguse avaldumist.

Kui kõik andmestikud olid koos, siis võeti ainult need vaatlused, kes olid jälgimise all olnud vähemalt 5 aastat või 5 aasta jooksul surnud.

Kuna KMI on metaboolse sündroomi määramisel oluline komponent, prioritseeriti, et see oleks igal indiviidil olemas. Jälgimisperioodi jooksul oli KMI mõõdetud 297-l indiviidil. Seetõttu lõplik juhtude valim moodustati psühhootilise diagnoosiga indiviididest, kellel oli KMI mõõdetud maksimaalselt kolm aastat enne psühhootilise haiguse avaldumist ning kellel ei esinenud enne seda metaboolse sündroomi tunnuseid.

Viimases etapis lisati samasse andmestikku tunnus, mis ütles, kas inimesel kujunes metaboolne sündroom viie aasta jooksul pärast haiguse avaldumist. 1633-st psühhootilise häire diagnoosiga indiviidist jäi lõplikku valimisse 220, kes vastasid kõigile tingimustele ning neist 40-l tekkis metaboolne sündroom 5

aasta jooksul pärast psühhootilise haiguse diagnoosi. Joonisel 1 on kujutatud, kuidas uuringusse kaasatud indiviidide valim andmestike põhjal saadi.



Joonis 1: Juhtude valimi saamine andmestike põhjal.

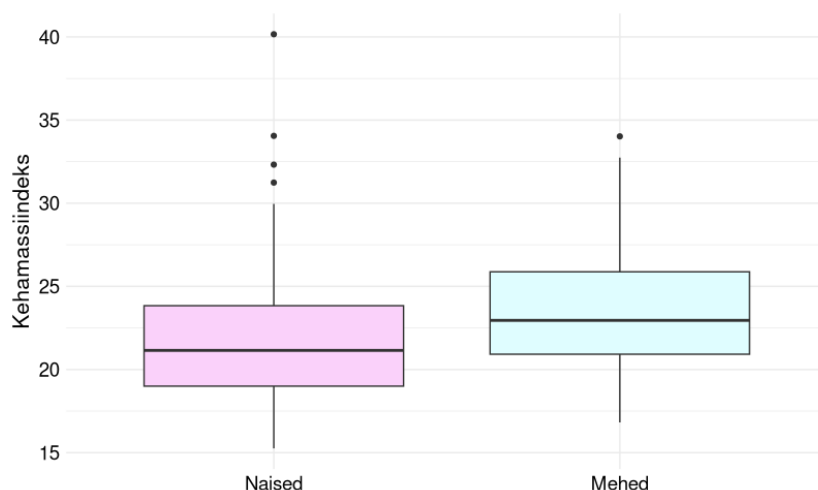
Järgnevalt on toodud näiteid indiviididest, kes olid jälgimise all olnud vähemalt 5 aastat või 5 aasta jooksul surnud ning kes kaasati lõplikku juhtude valimisse või jäeti sellest välja. Näiteks kui kahel indiviidil olid kõik vajalikud kriteeriumid täidetud, kuid ühel esines RHK-10 E11 kirje enne psühhootilise haiguse avaldumise kuupäeva ja teisel alles pärast seda, siis jäeti valimist välja indiviid, kellel E11 esines enne haiguse avaldumist. Teise näitena võib tuua olukorra, kus kahel inimesel olid TG, HDL, veresuhkru ja SBP mõõtmised normi sees kuni kolm aastat enne haiguse avaldumise kuupäeva, kuid ühel neist puudus samas ajavahemikus KMI mõõtmine. Siis kaasati lõplikku

valimisse inimene, kellel oli olemas ka KMI mõõtmine.

3.1.1 Kirjeldav analüüs

Lõplik andmestik koosnes 220 psühhootilise diagnoosiga indiviidist, kellest naisi oli 140 ja mehi 80. Naistest suitsetas või oli kunagi suitsetanud 55% ning meestest 61,2%. Juhtude valimis oli ka 2 naist, kes jälgimisperioodiks valitud 5 aasta jooksul surid.

Inimeste mediaanvanus psühhootilise haiguse avaldumisel oli 26,57 aastat. Noorim inimene oli 14,29 ja vanim 49,34. 75% vaatlustest olid vanuses kuni 33,5 aastat.

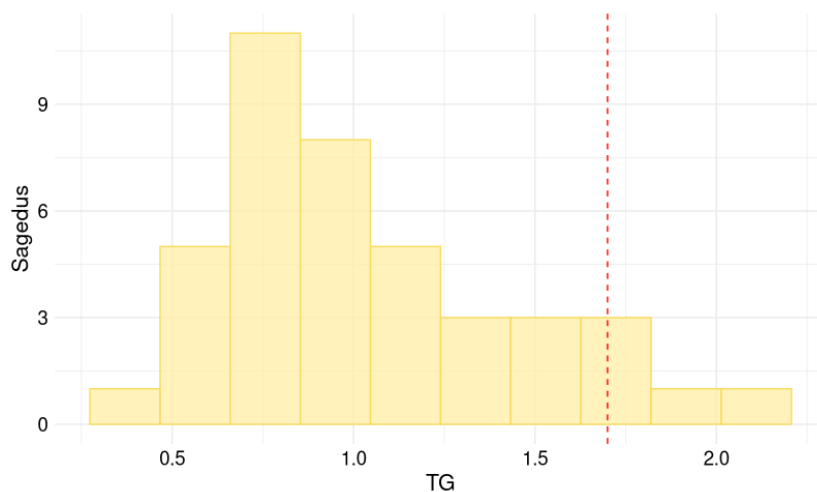


Joonis 2: Naiste ja meeste kehamassiindeksid juhtudel.

Karpdiagrammidel (vt joonis 2) on kujutatud naiste ja meeste KMI-d. Naistest pooltel jäi KMI vahemikku 19,0–23,8 ning mediaan oli 21,2. Meestest jäi pooltel KMI aga vahemikku 20,9–25,9 ning mediaan oli 23,0. Väikseim KMI oli naistel ja meestel vastavalt 15,3 ning 16,8, suurim 40,2 ja 34,0.

Triglütseriidide (TG) mõõtmistulemused olid olemas vaid 41-l psühhootilise

häire diagnoosiga indiviidil. Joonisel 3 on esitatud olemasolevate mõõtmistulemuste põhjal TG jaotus. Punase katkendliku joonega on kujutatud normi piir, millest suuremaid väärtusi loetakse normist kõrgemaks.



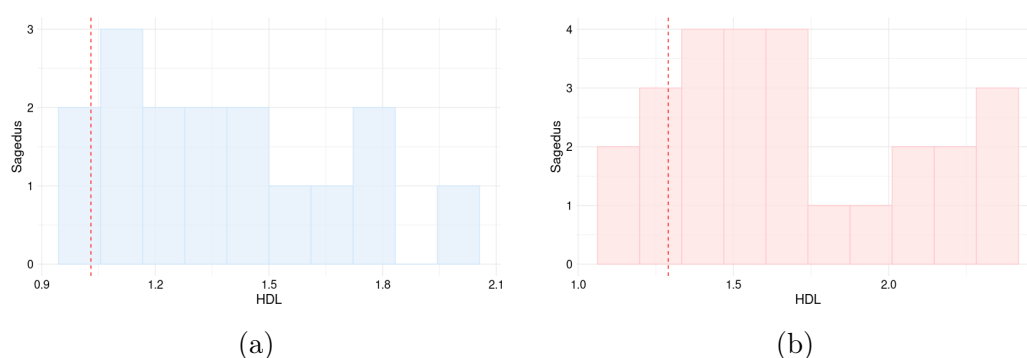
Joonis 3: Triglütseriidide jaotus.

Jaotus on parempoolse sabaga, mis tähendab, et enamik TG väärtusi on madalamad ning esineb üksikuid kõrgemaid väärtusi. Pooltel indiviididest jäi see vahemikku 0,77–1,3 mmol/l ning mediaan oli 0,98 mmol/l. Madalaim TG tase oli 0,37 mmol/l ja kõrgeim 2,11 mmol/l.

Lõplikus valimis oli glükoosi tase olemas 156 inimesel. Enamus nendest väärtustest jäi normaalsesse vahemikku, kuid leidis väärtusi, mis olid üle normi. Kõrgeim veresuhkru tase oli 13,3 mmol/l ning 25% väärtustest oli üle 5,7 mmol/l. Madalaim tase oli 3,4 mmol/l ning mediaan oli 5,01 mmol/l.

Kuna enamikul juhtude valimi indiviididest (196) oli olemas süstoolse vererõhu väärtus, võib selle põhjal esitatud kirjeldavaid statistilisi näitajaid pidada usaldusväärseks. Olemasolevate mõõtmistulemuste mediaan oli 120 mmHg ning keskmine 119,3 mmHg. Madalaim väärtus oli 90 mmHg ja kõrgeim 167 mmHg.

Joonisel 4 on kujutatud juhtude valimi meeste ja naiste HDL-kolesterooli tasemete sagedusi, kus punane katkendlik piirjoon tähistab normi piiri, millest väiksemaid väärtusi loetakse normist madalamaks. HDL väärtused olid kättesaadavad kokku 42 indiviidil, kellest 26 olid naised ja ülejäänud mehed, mistõttu ei ole statistilised näitajad niivõrd usaldusväärsed. Inimesi, kellel oli TG ja HDL väärtustest vähemalt üks, oli kokku 48.

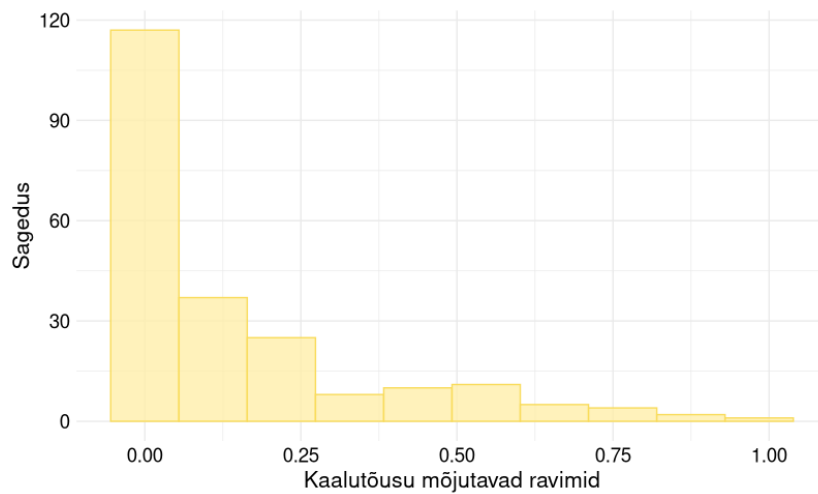


Joonis 4: HDL jaotus meeste (a) ja naiste (b) seas.

Naiste puhul jäi pooltel HDL-kolesterooli tase vahemikku 1,41 mmol/l–2,02 mmol/l ning mediaan oli 1,61 mmol/l. Meestest jäi pooltel HDL vahemikku 1,11 mmol/l–1,60 mmol/l ning mediaan oli 1,33 mmol/l. Madalaim HDL-kolesterooli tase oli naistel 1,13 mmol/l ja meestel 1,00 mmol/l, kõrgeim vastavalt 2,35 mmol/l ja 2,00 mmol/l.

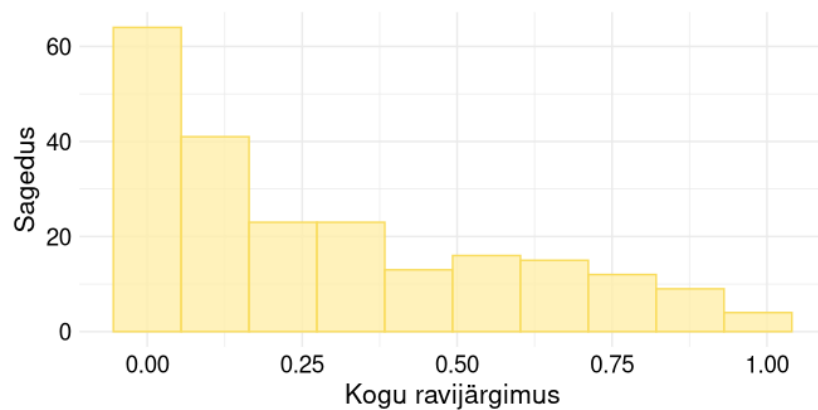
Kuna Eesti geenivaramus olid triglütseriidide ja HDL väärtused olemas vaid väga väheste inimeste jaoks, otsustati käesolevas töös valideerida vaid osalist mudelit, mis nimetatud näitajate väärtuseid ei kasutanud.

Järgmisel joonisel (vt joonis 5) on kujutatud juhtude valimis olevate individide kaalutõusu mõjutavate ravimite võtmise proportsiooni jälgimisperioodi jooksul.



Joonis 5: Kaalutõusu mõjutavate ravimite võtmise proportsiooni jaotus.

Näeme, et enamik psühhootilise diagnoosiga inimesi ei võtnud üldse kaalutõusu mõjutavaid ravimeid. Mediaan on 4,4% ning maksimaalselt võeti neid 98,4% jälgimisperioodi ajast.



Joonis 6: Antipsühhootiliste ravimite kasutamise proportsiooni jaotus.

Joonisel 6 on kujutatud indiviidide antipsühhootiliste ravimite kasutamise proportsiooni jaotust, mis näitab, kui suure osa jälgimisperioodist individ ravimit kasutas. Pooltel inimestel jäi ravimi võtmise proportsioon vahemikku 4,0%–50,1%. Mediaan oli 17,6% ning maksimaalne väärtus 98,5%.

3.2 Kontrollide andmestik

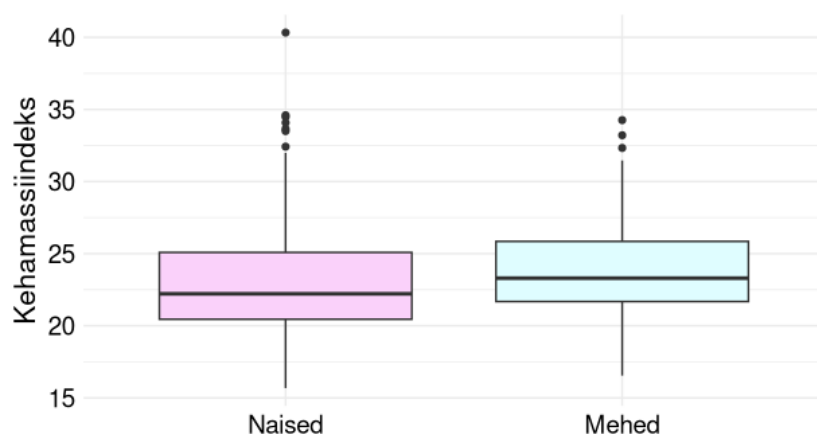
Kontrollide andmestik oli varasemalt defineeritud. See sisaldas igale psühhootilise häirega indiviidile kuni 3 kontrolli vastavalt eelpool nimetatud kriteeriumitele. Lõplik kontrollide andmestik filtreeriti nii, et vaatluse all oleksid ainult 220-le juhule vastavad kontrollid. Seega loeti lõplikuks 645 vaatlusega andmestik, kellest 46 tekkis hiljem metaboolne sündroom. Edasine andmetöötlus viidi läbi analoogselt juhtude andmestikule.

3.2.1 Kirjeldav analüüs

645 indiviidist koosnevast kontrollide valimist olid 412 naised ning 233 mehed. Naistest suitsetas või oli kunagi suitsetanud 34,5% ja meestest 48,9%.

Kontrollide valimis oli 5 aasta jooksul 3 surmajuhtu, kõik meessoost isikud.

Kontrollide vanus arvutati sünniaasta ja sobitatud alguskuupäeva aasta põhjal. Mediaanvanus oli 27 aastat. Noorim indiviid oli 14-aastane ja vanim 49-aastane. 75% vaatlustest olid vanuses kuni 34 aastat. Seega on kontrollide vanuse jaotus üsna sarnane juhtudele.



Joonis 7: Naiste ja meeste kehamassiindeksid kontrollidel.

Karpdiagmmidel (vt joonis 7) on kujutatud naiste ja meeste KMI-d. Naistest pooltel jäi KMI vahemikku 20,45–25,09 ning mediaan oli 22,21. Keskmise KMI näit oli 22,84. Meestest jäi pooltel KMI vahemikku 21,68–25,85. Mediaan oli 23,30 ja keskmine 23,83. Väikseim KMI oli naistel ja meestel vastavalt 15,68 ning 16,54, suurim 40,33 ja 34,26. Kuna nii meeste kui naiste puhul on keskmised ja mediaanid peaaegu samad, siis jaotus on ligikaudu sümmeetriline.

Lõplikus valimis oli glükoosi tase olemas 67-l inimesel. Enamus nendest väärtustest jäi normaalsesse vahemikku, kuid leidsid väärtusi, mis olid üle normi. Kõrgeim veresuhkru tase oli 11,5 mmol/l ning 25% väärtustest olid üle 5,3 mmol/l. Madalaim tase oli 4,2 mmol/l ning mediaan oli 4,9.

Süstoolse vererõhu mõõtmised soovitud ajavahemikus olid olemas 186-l indiviidil (ligikaudu 28,8%-l). Olemasolevate mõõtmistulemuste mediaan oli 118 mmHg ja keskmine 117,3 mmHg. Madalaim väärtus oli 75 mmHg ning kõrgeim 230 mmHg.

Kuna nii triglütseriidide kui ka HDL-kolesterooli mõõtmistulemuste väärtused olid kättesaadavad vaid 14-l inimesel (ligikaudu 2,2%-l), ei ole nende jaotused piisavalt esinduslikud, et teha usaldusväärseid järeldusi. Seetõttu nende jaotusi käesolevas töös ei kujutatud.

4 Analüüsitulemused

Järgnev peatükk annab ülevaate Eesti geenivaramu andmete analüüsist, analüüsitulemustest ja järeldustest.

4.1 Mudeli testimine juhtudel

Saadud juhtude andmestikku kasutati osalise mudeli rakendamiseks (vt tabel 2). Kuna Eesti geenivaramu andmestikku, mille põhjal käesolevas töös analüüsid teostati, kuulusid vaid Euroopa päritolu inimesed, olid päritolu näitavate indikaatortunnuste väärtused kõigi testandmestikku kuuluvate inimeste jaoks nullid. Seetõttu ei andnud need tunnused selles andmestikus mudeli prognoosidesse panust, mistõttu järgnevates mudelites pole neid tunnuseid ja nendele vastavaid kordajaid välja kirjutatud. Samuti polnud antipsühhootiliste ravimite kasutus kategooriline tunnus, vaid pidev, s.t arvutatud ravimikasutuse osakaaluna jälgimisaja jooksul (osakaalu arvutamise kirjeldus on alapeatükis 3.1). Seetõttu hinnati kahte erinevat osalist mudelit: ühes kasutati tunnust, mis kirjeldab kaalutõusu mõjutavate ravimite kasutust, ning teises kaasati tunnus, mis kirjeldab kogu ravi järgimusest vaatlusaja jooksul. Vastavad mudelid olid kujul:

$$\begin{aligned}\text{logit}(p) = & -6,973829 + 0,00633115 \cdot \textit{vanus} + 0,16912161 \cdot \textit{KMI} \\ & + 0,31460036 \cdot \text{I}(\textit{sugu} = \text{„mees“}) \\ & + 0,24751854 \cdot \text{I}(\textit{suitsetamise_staat} = \text{„suitsetanud kunagi/praegu“}) \\ & + 0,60013558 \cdot \textit{kaalutõusu_mõjutavad_ravimid}\end{aligned}$$

$$\begin{aligned}\text{logit}(p) = & -6,973829 + 0,00633115 \cdot \textit{vanus} + 0,16912161 \cdot \textit{KMI} \\ & + 0,31460036 \cdot \text{I}(\textit{sugu} = \text{„mees“}) \\ & + 0,24751854 \cdot \text{I}(\textit{suitsetamise_staatus} = \text{„suitsetanud kunagi/praegu“}) \\ & + 0,60013558 \cdot \textit{kogu_ravi_järgimus},\end{aligned}$$

kus p tähistab metaboolse sündroomi tekkimise tõenäosust jälgimisperioodi jooksul. Algses uuringus prognoosis mudel metaboolse sündroomi riski kuni kuue aasta jooksul (Perry *et al.*, 2022), kuid käesolevas töös rakendatakse mudelit viieaastase jälgimisperioodi põhjal koostatud andmestikul.

Mõlema mudeli puhul leidis Hosmer-Lemeshow test, et mudel ei sobi vaatlusandmetega ($p < \alpha = 0,05$). See võib olla tingitud sellest, et esialgne mudel hinnati Ühendkuningriigi andmete peal ning seega ei pruugi mudeliga saadud hinnangud olla Eesti andmestikul sama hästi kalibreeritud. Põhjuseks võivad olla näiteks populatsioonidevahelised erinevused tervishoius ja elustiilis.

Seetõttu hinnati uus logistilise regressiooni mudel, kasutades eelnevate mudelite lineaarset prediktorit ainsa seletava tunnusena. Mõlema mudeli korral osutus lineaarne prediktor statistiliselt oluliseks ning Hosmer-Lemeshow testi põhjal ei olnud alust nullhüpoteesi tagasi lükata (esimese mudeli puhul $p = 0,8517$ ja teise puhul $p = 0,5546$). Seega ei ole testi põhjal alust järeldada, et mudel ei sobi andmetega. Mudelite võrdlemisel selgus, et mudelil, milles kasutati kaalutõusu mõjutavate ravimite tunnust, oli veidi suurem AIC (181,96) ja väiksem AUC (0,753) kui mudelil, milles kasutati kogu ravi järgimuse tunnust (AIC = 181,27; AUC = 0,759). Erinevused olid nii AIC kui ka AUC väärtustes küll väikesed, kuid mõlemad viitasid sellele, et viimane mudel on parem. Seetõttu võib kõigi antipsühhootiliste ravimite kasutust

arvestavat mudelit pidada paremaks. Tabelis 5 on selle mudeli parameetrid koos standardvea ja Wald-testi p-väärtusega.

Tunnus	β	SD	p-väärtus
Vabaliige	1,87	0,66	0,005
Lp	1,42	0,28	$4,86 \cdot 10^{-7}$

Tabel 5: Rekalibreerimismudel.

Nagu eelnevalt mainitud, oli logistilise regressiooni mudeli, mille ainsaks seletavaks tunnuseks oli kogu ravi järgimust sisaldava mudeli lineaarne prediktor, AUC 0,759. Algse osalise mudeli valideerimisanalüüsis ALSPAC valimil oli AUC 0,71 (Perry *et al.*, 2021), seega on tulemused üsna sarnased.

Logistilise regressiooni mudeliga, mille seletavaks tunnuseks oli kogu ravi järgimusega mudeli lineaarne prediktor, saavutati ROC-kõvera põhjal tundlikkus 67,5% ja spetsiifilisus 77,8%, kui metaboolse sündroomi esinemiseks loeti juhud, mille korral hinnatud tõenäosus ületas piiri 0,2016084. Piirväärtus määrati ROC-analüüsi põhjal optimaalse lävendina nii, et tundlikkuse ja spetsiifilisuse summa oleks maksimaalne. Tundlikkus 67,5% tähendab, et 67,5% neist psühhootilise diagnoosiga indiviididest, kellel tegelikult viie aasta jooksul kujunes välja metaboolne sündroom, oli kasutatud mudeli (ja piirväärtuse) poolt lahterdatud kõrgeenenud metaboolse riskiga patsientideks. Spetsiifilisus 77,8% tähendab, et indiviididest, kellel metaboolset sündroomi ei diagnoositud järgneva 5 aasta jooksul, oli 77,8% lahterdatud madala riskiga patsientide sekka.

Järgmises tabelis (vt tabel 6) on mudeli prognooside ja tegelike väärtuste sagedused, kust on näha, kui paljudel juhtudel andis mudel metaboolse sündroomi esinemise õigesti ning kui paljudel juhtudel eksis.

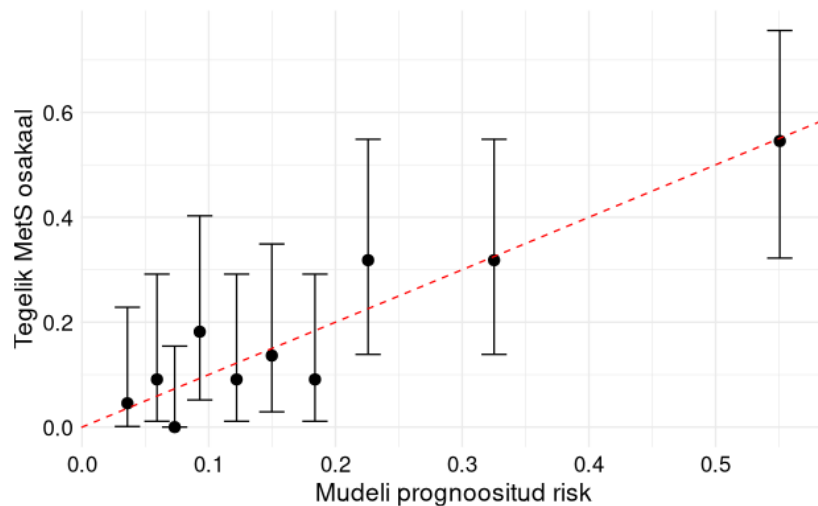
Prognoos	Tegelik väärtus		Kokku
	0	1	
0	140	13	153
1	40	27	67
Kokku	180	40	220

Tabel 6: Mudeli prognooside ja tegelike väärtuste võrdlus piirmäära 0,2016084 kasutades. Ühega on tähistatud individid, kellel tekkis metaboolne sündroom järgneva 5 aasta jooksul ning nulliga need, kellel metaboolset sündroomi ei avaldunud.

Mudel leidis korrektselt üles 27 indiviidi 40-st, kellel diagnoositi järgneva 5 aasta jooksul metaboolne sündroom (tähistatud ühega). 180-st inimesest, kellel ei avaldunud jälgimisperioodi jooksul metaboolset sündroomi (tähistatud nulliga), lahterdas mudel korrektselt 140 indiviidi. Kuna mudel lahterdas aga 40 indiviidi, kellel metaboolset sündroomi järgneva 5 aasta jooksul ei tekkinud, ekslikult nende hulka, kellel sündroom tekkis, siis võib öelda, et mudel kaldub andma suhteliselt palju valepositiivseid tulemusi. Kuna aga metaboolse sündroomi varajane avastamine ja ennetamine on oluline, siis selline mudel võib olla põhjendatud.

Järgmisel joonisel (vt joonis 8) on kujutatud kalibreeritud mudeli prognoositud riskid ja tegelikud metaboolse sündroomi osakaalud. Mudeliga saadud hinnatud tõenäosused on jaotatud kümneks grupiks, milles on enam-vähem võrdne arv vaatlusi. Nendes gruppides arvutati binoomtesti abil tegelike tõenäosuste 95% usaldusintervalli ülemine ja alumine piir. Punane katkendlik joon kujutab ideaalset olukorda, kus hinnatud tõenäosused vastaksid tegelikele väärtustele. Kuna punktid ja nende 95% usaldusintervallid on üsna lähedal diagonaalile ning jälgivad üldist trendi, siis võib joonise põhjal järeldada, et mudeli prognoositud tõenäosused on heas kooskõlas tegelike esinemissage-

dustega.



Joonis 8: Kalibreeritud mudeli prognoositud tõenäosuste ja metaboolse sündroomi (MetS) tegeliku esinemisosakaalu võrdlus detsiilide kaupa.

4.2 Mudeli testimine kontrollrühmas

Kuna üheks eesmärgiks oli uurida, kas psühhootilise diagnoosiga indiviididel peaks rakendama üldpopulatsioonist erinevat metaboolse sündroomi riski prognoosivat mudelit, võeti vaatluse alla logistiline regressiooni mudel, mille ainsaks seletavaks tunnuseks oli kogu ravi järgimusega mudeli lineaarne prediktor ja testiti seda kontrollindiviididel. Kuna kontrollrühmas ravimeid ei võetud, oli kogu ravi järgimusega tunnuse väärtused nende puhul nullid ning see tunnus ei mõjutanud kontrollindiviidide lineaarse prediktori väärtust.

Kontrollrühmas arvutati tundlikkus ja spetsiifilisus, kasutades sama otsustustälvendit, mis leiti psühhootilise diagnoosiga indiviididel, et hinnata mudeli üldistusvõimet. Tundlikkus tuli seega 58,7% ja spetsiifilisus 77,8%. Kuigi spetsiifilisus oli nii kontrollidel kui ka juhtudel sama, siis tundlikkus erines üsna palju. Väiksem tundlikkus kontrollidel viitab sellele, et mudel tuvas-

tab metaboolse sündroomiga inimesi neil halvemini. Samas AUC (0,758) on peaaegu sama ehk mudel eristab hästi, kuid ei anna õigeid tõenäosuseid.

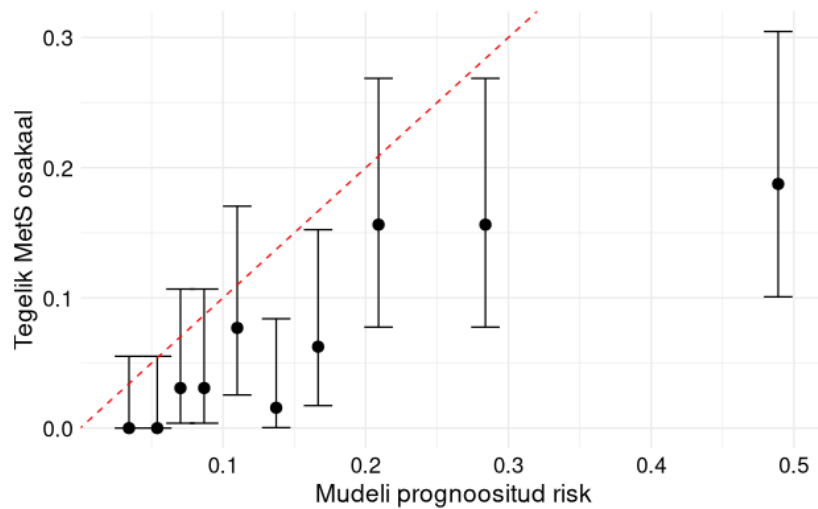
Järgmises tabelis (vt tabel 7) on mudeli prognooside ja tegelike väärtuste sagedus, kus hinnatud tõenäosuse piiriks loeti jälle 0,2016084.

Prognoos	Tegelik väärtus		Kokku
	0	1	
0	466	19	485
1	133	27	160
Kokku	599	46	645

Tabel 7: Mudeli prognooside ja tegelike väärtuste võrdlus piirmäära 0,2016084 kasutades. Ühega on tähistatud individid, kellel tekkis metaboolne sündroom järgneva 5 aasta jooksul ning nulliga need, kellel metaboolset sündroomi ei avaldunud.

46-st kontrollist, kellel diagnoositi järgneva 5 aasta jooksul metaboolne sündroom (tähistatud ühega), leidis mudel korrektselt üles 27. 599-st inimesest, kellel ei avaldunud jälgimisperioodi jooksul metaboolset sündroomi (tähistatud nulliga), lahterdas mudel korrektselt 466 indiviidi. Mudel lahterdas ekslikult 133 indiviidi nende hulka, kellel kujunes järgneva viie aasta jooksul metaboolne sündroom.

Joonisel 9 on kujutatud psühhootilise diagnoosiga individide kalibreeritud mudeliga saadud hinnatud tõenäosusi kontrollidel ja nende tegelikke metaboolse sündroomi osakaale koos 95% usaldusintervalliga kümnes grupis, mis olid tehtud sarnaselt psühhootilise diagnoosiga individide joonisele.



Joonis 9: Kalibreeritud mudeli prognoositud tõenäosuste ja metaboolse sündroomi (MetS) tegeliku esinemisosakaalu võrdlus detšiilide kaupa kontrollidel.

Jooniselt on näha, et mudeliga hinnatud tõenäosused ei vasta eriti hästi tegelikkusele. Kuna punktid paiknevad allpool diagonaali, viitab see sellele, et mudel hindab metaboolse sündroomi tekkimise riski üle. Ka Hosmer-Lemeshow test andis tulemuseks, et mudel ei sobi andmetega ($p < 0,05$). Seega võib järeldada, et antud mudel ei sobi metaboolse sündroomi riski hindamiseks vaatluse all olevatele kontrollidele ega pruugi sobida ka üldpopulatsioonile.

Järgmiseks liideti juhtude ja kontrollide andmestikud kokku, milles defineeriti, kumma andmestikuga on tegu ning lisati vastavalt eelnevalt tuletatud lineaarse prediktori tunnus mudeli põhjal, kus üheks seletavaks tunnuseks oli kogu ravi järgimuse tunnus. Seejärel hinnati andmestiku põhjal kaks logistilise regressiooni mudelit: üks, mille seletavateks tunnusteks olid lineaarne prediktor, kontrolle ja juhte eristav tunnus ning nende kahe tunnuse koosmõju, ja teine, mille seletavaks tunnuseks oli ainult lineaarne prediktor. Mudelite võrdlemiseks kasutati tõepärasuhte testi. Tulemuseks saadi, et ainult

lineaarse prediktoriga mudel erines statistiliselt oluliselt koosmõjudega mudelist ($p = 0,0001424 < 0,05$). Kuna seos prognoositud riski ja metaboolse sündroomi tegeliku esinemise vahel erineb kontrollidest, võib järeldada, et psühhootilise diagnoosiga indiviidide puhul tasub kasutada spetsiaalselt nende jaoks välja töötatud mudelit. See tulemus näitab, et psühhootilise häire diagnoosiga ravi vajavate inimeste risk saada järgneva 5 aasta jooksul metaboolse sündroomi diagnoos, ka samade seletavate tunnuste väärtuste korral, on suurem kui üldpopulatsiooni kuuluvatel isikutel.

4.3 Mudeli hindamine juhtudel

Lisaks eelnevale, hinnati psühhootilise diagnoosiga indiviididele ka kaks uut logistilise regressiooni mudelit, mis erinesid ravi järgimuse tunnuse poolest. Mõlema mudeli algses versioonis olid seletavateks tunnusteks KMI, tunnus, mis ütles, kas indiviid on või ei ole kunagi suitsetanud, vanus haiguse avaldumise hetkel, sugu, TG, HDL, veresuhkur, SBP ning vastavalt proportsioon kaalutõusuga seotud ravimite või siis kogu ravi järgimuse kohta jälgimisaja jooksul. Seejärel jäeti seletavad tunnused ükshaaval välja, alustades suurima p -väärtusega ja statistiliselt mitteolulisest tunnusest. Lõpuks saadi kaks mudelit, kus statistiliselt olulisteks ($p < 0,05$) tunnusteks olid KMI, veresuhkru tase ja kaalutõusuga seotud ravimite järgimus või kogu ravi järgimus. Kuna HDL-i ja TG puhul oli andmestikus palju puuduvaid väärtusi ja need asendati keskmisega, vähenes nende tunnuste varieeruvus. Tõenäoliselt seetõttu ei osutunud need mudelis metaboolse sündroomi riski prognoosimisel statistiliselt olulisteks. Hosmer-Lemeshow testi põhjal sobisid mõlemad mudelid andmetega ($p > 0,05$).

Kuna mudelite AIC ja AUC erinevused olid väga väikesed (AIC vastavalt

177,3 ja 176,4 ning AUC 0,777 ja 0,775) ning kumbki mudel ei osutunud mõlema näitaja põhjal selgelt paremaks, siis ei olnud vahet, kumba mudelit kasutada. Kogu ravi järgimusega mudelil oli AUC 0,777 ning kaalutõusuga seotud ravimitega mudelil oli see 0,775 ehk mõlemal mudelil oli küllalt hea eristusvõime. Tundlikkus ja spetsiifilisus olid kogu ravi järgimusega mudelil veidi paremad: kogu ravi järgimusega mudelis olid need vastavalt 70% ja 77,8% ning kaalutõusuga seotud ravimitega mudelis vastavalt 67,5% ja 76,7%. Nimetatud tulemustest lähtuvalt ja toetudes teadmisele, et kogu ravi järgimus hõlmab ka kaalutõusuga seotud ravimite võtmist, otsustati kogu ravi järgimust sisaldava mudeli kasuks. Järgmises tabelis (vt tabel 8) on mudeli parameetrid koos standardvea, tõepärasuhte testi p-väärtuse, šansside suhte ning selle 95% usaldusintervalliga.

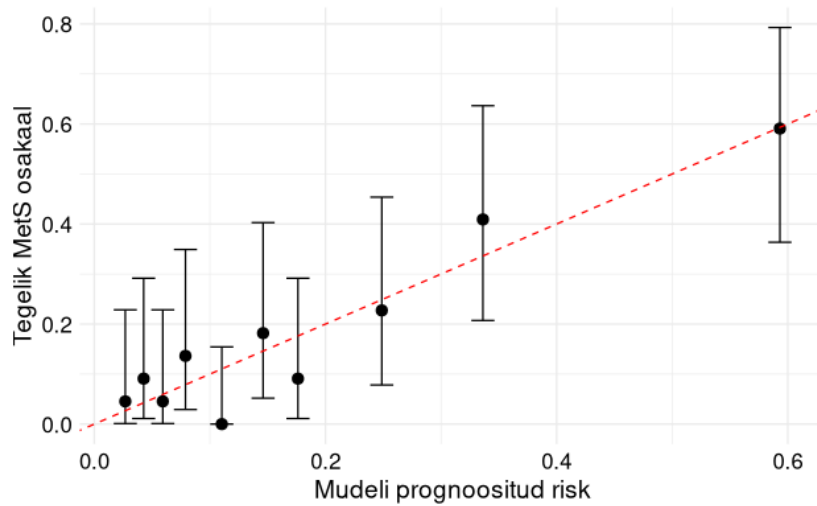
Tunnus	β	SD	OR	95% CI	p-väärtus
Vabaliige	-10,42	1,85	–	–	–
KMI	0,27	0,05	1,32	(1,19; 1,47)	$1,097 \cdot 10^{-8}$
Veresuhkru tase	0,37	0,20	1,45	(1,01; 2,20)	0,04
Kogu ravi järgimus	1,68	0,68	5,38	(1,43; 20,64)	0,01

Tabel 8: Logistilise regressiooni mudel.

Kui KMI suureneb ühe ühiku võrra, suureneb metaboolse sündroomi kujunemise šanss 1,32 korda. Veresuhkru taseme suurenemisel ühe mmol/l võrra on metaboolse sündroomi tekkimise šanss 45% suurem. Kuigi hinnangu standardviga on suhteliselt suur, ei sisalda šansi 95% usaldusvahemik väärtust üks, mistõttu loetakse seos statistiliselt oluliseks. Ravi järgimuse ühe ühiku suurenemisel on metaboolse sündroomi kujunemise šanss 5,38 korda suurem. 95% usaldusintervall on lai (1,43–20,64), viidates seega hinnangu ebakindlusele. See võib olla tingitud väikesest valimimahust või suurest varieeruvusest andmetes.

Joonisel 10 on näha, kui hästi mudel prognoosib metaboolse sündroomi ku-

junemise riski võrreldes tegelikkusega.



Joonis 10: Juhtude andmestikul hinnatud mudeli prognoositud tõenäosuste ja metaboolse sündroomi (MetS) tegeliku esinemisosakaalu võrdlus detsiilide kaupa.

Mudel töötab paremini kõrgemate riskide korral. Mõnes kohas prognoosib mudel madalate riskide puhul tegelikkusest suuremaid riske. Ka usaldusintervallid on laiad ehk järeldusi tuleb teha ettevaatusega.

Järgmiseks tekitati sarnaselt eelnevale koondandmestik juhtude ja kontrollide andmestike kokkuliitmisel, kuhu lisati tunnus, mis eristas juhte ja kontrole. Samuti lisati kõigile vaatlustele lineaarne prediktor, kasutades eelnevalt juhtude andmestikul hinnatud mudeli kordajaid. Seejärel hinnati andmestiku põhjal kaks logistilise regressiooni mudelit: ühes oli seletavateks tunnusteks lineaarne prediktor, kontrole ja juhte eristav tunnus ning nende kahe tunnuse koosmõju, ja teises oli seletavaks tunnuseks ainult lineaarne prediktor. Koosmõjuga mudelis ei osutunud koosmõju statistiliselt oluliseks ($p = 0,26 > 0,05$). Mudelite võrdlemisel tõepärasuhte testiga selgus, et ainult lineaarse prediktoriga mudel erines statistiliselt oluliselt koosmõjudega

modelist ($p = 0,004 < 0,05$), mistõttu võib taaskord järeldada, et psühhootilise diagnoosiga indiviidide puhul võib neile sobitatud mudeli kasutamine olla vajalik. Seda kinnitas ka see, et seos prognoositud riski ja metaboolse sündroomi tegeliku esinemise vahel erineb kontrollidest.

4.4 Rek calibreeritud mudeli ja juhtudel hinnatud mudeli võrdlus

Tabelis 9 on juhtudel hinnatud mudeli ja juhtudel kõigi antipsühhootiliste ravimite kasutust arvestava rek calibreeritud mudeli AIC, AUC, tundlikkus ja spetsiifilisus.

Mudel	AIC	AUC	Tundlikkus	Spetsiifilisus
Rekalibreeritud mudel	181,27	0,759	67,5%	77,8%
Juhtudel hinnatud mudel	177,33	0,777	70%	77,8%

Tabel 9: Juhtude andmestikul hinnatud mudelite võrdlus.

Andmetel hinnatud mudeli AUC ja tundlikkus olid suuremad ning AIC väiksem kui osalise mudeli lineaarset prediktorit kasutaval kalibreeritud mudelil. See viitab, et juhtudel hinnatud mudel on veidi parem metaboolse sündroomi riski prognoosimisel psühhootilise diagnoosiga inimestel. Samas, kuna erinevused on siiski küllalt väikesed, ei saa ühte mudelit pidada selgelt paremaks. Tulemuste põhjal võib samuti järeldada, et Ühendkuningriigis väljatöötatud osalist mudelit on võimalik kasutada ka Eesti andmetel, kuid selle kasutamine eeldab rek calibreerimist.

Kokkuvõte

Bakalaureusetöö eesmärk oli testida ja vajadusel modifitseerida Tartu Ülikooli Eesti geenivaramu andmete peal Ühendkuningriigis arendatud mudeleid, millega on võimalik prognoosida psühhootilise häire diagnoosiga indiviididel metaboolse sündroomi kujunemise riski järgmise viie aasta jooksul. Samuti oli üheks eesmärgiks hinnata, kas psühhootilise diagnoosiga indiviidide jaoks koostatud metaboolse sündroomi riski prognoosiv mudel on sobiv kasutamiseks ka indiviididel, kellel psühhootilist häiret ei esine.

Ühendkuningriigis oli välja arendatud nii osaline kui ka täismudel. Täismudel sisaldas lisaks teistele tunnustele ka triglütseriide ja HDL-kolesterooli näitu. (Perry *et al.*, 2022) Kuna Eesti geenivaramus oli nimetatud tunnuste mõõtmistulemuste väärtused vastavas vaatlusaknas puudulikud, otsustati käesolevas töös viia analüüs läbi vaid osalise mudeliga. Vaatlusandmetest lähtuvalt võeti Eesti geenivaramu andmetes vaatluse alla viie aasta pikkune periood.

Kuigi vaatluse all olev mudel sisaldas binaarset tunnust metaboolselt aktiivsete ehk kaalutõusu mõjutavate antipsühhootikumide tarvitamise kohta vastavalt jälgimisaja alguses, kasutasime meie oma analüüsis nimetatud tunnuse asemel kahte modifitseeritud versiooni: kaalutõusu mõjutavate ravimite või kõikide antipsühhootikumide või meeleolu stabilisaatorite kasutamise proportsiooni jälgimisaja jooksul. Hosmer-Lemeshow testi tulemusena selgus, et mudelid ei sobinud andmetega hästi ($p < 0,05$), mistõttu hinnati kaks uut logistilise regressiooni mudelit, kasutades eelnevate mudelite lineaarset prediktorit ainsa seletava tunnusena. Need mudelid sobisid nii Hosmer-Lemeshow testi kui ka muude näitude põhjal (AUC, spetsiifilisus, tundlikkus), kuid parimaks osutus kogu ravi järgimust kirjeldavat tunnust sisaldanud kalibreeritud mudel. Mudeliga saavutati tundlikkus 67,5% ja spetsiifilisus 77,8%, kui me-

taboolse sündroomi esinemiseks loeti juhud, mille korral hinnatud tõenäosus ületas piiri 0,2016084. Antud mudeli testimisel kontrollrühmas selgus, et mudel ei sobi metaboolse sündroomi riski hindamiseks indiviididel, kellel psühhootilist häiret ei esinenud ja seega ei pruugi sobida üldpopulatsioonile.

Lisaks olemasoleva mudeli testimisele hinnati Eesti geenivaramu andmetel ka uus logistilise regressiooni mudel. Lõplik mudel sisaldas kolme statistiliselt olulist tunnust ($p < 0,05$): glükoosi tase, kehamassiindeks ja kogu ravi järgimus. Hosmer-Lemeshow testi põhjal sobis mudel andmetega ($p > 0,05$). Mudel töötas paremini kõrgemate riskide korral. Ka selle analüüsi põhjal sai järeldada, et psühhootilise diagnoosiga indiviididele võiks olla üldpopulatsioonist erinev metaboolse sündroomi riski ennustamise mudel.

Kokkuvõttes selgus, et Ühendkuningriigis väljatöötatud osalist mudelit on võimalik kasutada Eesti andmetel, kuid selle rakendamine eeldab rekaliibreerimist. Kuna varasemalt on mudelit valideeritud ka Hispaania ja Šveitsi andmetel (Perry *et al.*, 2022) ning käesolevas töös ilmnes mudeli potentsiaal ka Eesti andmetel, viitavad tulemused mudeli võimalikule ülekantavusele teistesse Euroopa populatsioonidesse, kuid tõenäoliselt eeldab see samuti rekaliibreerimist. Lisaks toetasid tulemused arusaama, et psühhootilise häire diagnoosiga indiviididel võiks olla üldpopulatsioonist erinev metaboolse sündroomi riski prognoosiv mudel.

Kasutatud allikad

- Alver, Maris, Silva Kasela, Liina Haring, Laura Birgit Luitva, Estonian Biobank Research Team, Health Informatics Research Team, Krista Fischer, Märt Möls ja Lili Milani (2024). “Genetic predisposition and antipsychotic treatment effect on metabolic syndrome in schizophrenia: a ten-year follow-up study using the Estonian Biobank”. *The Lancet Regional Health* 41. DOI: [10.1016/j.lanepe.2024.100914](https://doi.org/10.1016/j.lanepe.2024.100914).
- Grande, Iria, Michael Berk, Boris Birmaher ja Eduard Vieta (2016). “Bipolar disorder”. *The Lancet* 387.10027, lk. 1561–1572. DOI: [10.1016/S0140-6736\(15\)00241-X](https://doi.org/10.1016/S0140-6736(15)00241-X).
- Hosmer, David W., Stanley Lemeshow ja Rodney X. Sturdivant (2013). *Applied Logistic Regression*. 3. väljaanne. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc. ISBN: 9780470582473.
- Jauhar, Sameer, Mandy Johnstone ja Peter J McKenna (2022). “Schizophrenia”. *The Lancet* 399.10323, lk. 473–486. DOI: [10.1016/S0140-6736\(21\)01730-X](https://doi.org/10.1016/S0140-6736(21)01730-X).
- Lim, Debbie Xiu En, Shi Yun Yeo, Zhen You Ashley Chia, Aaron Zefrin Ferdinandis, Jimmy Lee ja John Jia En Chua (2026). “Schizophrenia: Genetics, neurological mechanisms, and therapeutic approaches”. *Neural Regeneration Research* 21.3, lk. 1089–1103. DOI: [10.4103/NRR.NRR-D-24-01375](https://doi.org/10.4103/NRR.NRR-D-24-01375).
- Ma, Wen-Ya, Hung-Yuan Li, Chi Sheng Hung, Mao-Shin Lin, Fu-Chun Chiu, Cheng-Hsin Lin, Shyang-Rong Shih, Lee-Ming Chuang ja Jung-Nan Wei (2009). “Metabolic syndrome defined by IDF and AHA/NHLBI correlates better to carotid intima-media thickness than that defined by NCEP ATP III and WHO”. *Diabetes Research and Clinical Practice* 85.3, lk. 335–341. DOI: [10.1016/j.diabres.2009.06.020](https://doi.org/10.1016/j.diabres.2009.06.020).

- McIntyre, Roger S, Michael Berk, Elisa Brietzke, Benjamin I Goldstein, Carlos López-Jaramillo, Lars Vedel Kessing, Gin S Malhi, Andrew A Nierenberg, Joshua D Rosenblat, Amna Majeed, Eduard Vieta, Maj Vinberg, Allan H Young ja Rodrigo B Mansur (2020). “Bipolar disorders”. *The Lancet* 396.10265, lk. 1841–1856. DOI: [10.1016/S0140-6736\(20\)31544-0](https://doi.org/10.1016/S0140-6736(20)31544-0).
- OpenAI (2026). *ChatGPT [suur keelemudel]*. URL: <https://chat.openai.com/>.
- Perry, Benjamin I., Emanuele F. Osimo, Rachel Upthegrove, Pavan K. Mallikarjun, Jessica Yorke, Jan Stochl, Jesus Perez, Stan Zammit, Oliver Howes, Peter B. Jones ja Golam M. Khandaker (2021). “Development and external validation of the Psychosis Metabolic Risk Calculator (PsyMetRiC): a cardiometabolic risk prediction algorithm for young people with psychosis”. *The Lancet Psychiatry* 8.7, lk. 589–598. DOI: [10.1016/S2215-0366\(21\)00114-0](https://doi.org/10.1016/S2215-0366(21)00114-0).
- Perry, Benjamin I., Frederik Vandenberghe, Nathalia Garrido-Torres, Emanuele F. Osimo, Marianna Piras, Javier Vazquez-Bourgon, Rachel Upthegrove, Claire Grosu, Victor Ortiz-Garcia De La Foz, Peter B. Jones, Nermine Laaboub, Miguel Ruiz-Veguilla, Jan Stochl, Celine Dubath, Manuel Canal-Rivero, Pavan Mallikarjun, Aurélie Delacrétaz, Nicolas Ansermot, Emilio Fernandez-Egea, Severine Crettol, Franziska Gamma, Kerstin J. Plessen, Philippe Conus, Golam M. Khandaker, Graham K. Murray, Chin B. Eap ja Benedicto Crespo-Facorro (2022). “The psychosis metabolic risk calculator (PsyMetRiC) for young people with psychosis: International external validation and site-specific recalibration in two independent European samples”. *The Lancet Regional Health - Europe* 22. DOI: [10.1016/j.lanepe.2022.100493](https://doi.org/10.1016/j.lanepe.2022.100493).

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Birthe Lall,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose Psühhootilise häirega patsientidele mõeldud metaboolse sündroomi riski prognoosimudeli valideerimine Eesti geenivaramu andmetel, mille juhendajad on Maris Alver ja Märt Möls, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Birthe Lall

12.05.2026