

TARTU ÜLIKOOL

LOODUS- JA TÄPPISTEADUSTE VALDKOND

MATEMAATIKA JA STATISTIKA INSTITUUT

Saskia Kuusk

**Sihtkatse matkimine: polügeense riskiskoori ja
kolesterooli alandava ravi mõju
südame-veresoonkonna haiguste riskile**

Matemaatika ja statistika õppekava

Matemaatilise statistika eriala

Magistritöö (30 EAP)

Juhendajad: Prof. Krista Fischer, PhD

Prof. Lili Milani, PhD

TARTU 2024

**SIHTKATSE MATKIMINE: POLÜGEENSE RISKISKOORI JA
KOLESTEROOLI ALANDAVA RAVI MÕJU
SÜDAME-VERESONKONNA HAIGUSTE RISKILE**

Magistritöö

Saskia Kuusk

Lühikokkuvõte

Käesoleva magistritöö eesmärk on tutvustada sihtkatse matkimise ideed ja teostada praktiline näide. Selleks kasutatakse Tartu Ülikooli Eesti geenivaramu andmeid, et hinnata polügeense riskiskoori ja kolesterooli alandava ravi mõju vereringeelundite haiguste riskile. Pannakse paika hüpoteetiline sihtkatse, mida soovitakse vaatlusandmete pealt matkida koos vastavate protokollide komponentidega. Elukestusanalüüsi meetodite abil leiti, et kolesterooli alandavate ravimite määramine on seotud vähenenud müokardi- ja peaajuinfarktide ning kardiovaskulaarsete surma riskiga. Samuti valideeriti südameveresoonkonna haiguste riskiskoori mõju huvipakkuvate kardiovaskulaarsete sündmuste riskile, välja arvatud ainult kardiovaskulaarset surma käsitleva sündmuse puhul.

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

Märksõnad: Põhjuslikkus, suurandmed, kardiovaskulaarne risk, elukestusanalüüs, kolesteroolisisaldust vähendavad ained, farmakogeneetika.

**TARGET TRIAL EMULATION: THE EFFECT OF POLYGENIC
RISK SCORE AND CHOLESTEROL-LOWERING THERAPY ON
CARDIOVASCULAR DISEASE RISK**

Master thesis

Saskia Kuusk

Abstract

The aim of this master's thesis is to introduce the concept of target trial emulation and provide a practical example. This is achieved using data from the Estonian Biobank at the University of Tartu to evaluate the effect of polygenic risk scores and cholesterol-lowering treatment on the risk of cardiovascular diseases. A hypothetical target trial is established to be emulated using observational data with the corresponding protocol components. Using survival analysis methods, it was found that the prescription of cholesterol-lowering drugs is associated with a reduced risk of myocardial infarction, stroke, and cardiovascular death. Additionally, the impact of the cardiovascular disease risk score on the risk of relevant cardiovascular events was validated, except for the event concerning only cardiovascular death.

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

Key Words: Causality, big data, cardiovascular risk, survival analysis, cholesterol inhibitors, pharmacogenetics.

Sisukord

Sissejuhatus	4
1 Sihtkatse matkimine vaatlusandmete põhjal	5
1.1 Randomiseeritud katse	5
1.1.1 Põhjuslik seos	6
1.1.2 Randomiseerimine	7
1.1.3 Randomiseeritud katse komponendid	8
1.2 Vaatluuuringud ja kehtivate põhjuslike järelduste tingimused	10
1.3 Sihtkatse matkimine	11
1.3.1 Ülesehitus	11
1.3.2 Võimalikud nihkeallikad	13
1.3.3 Järjestikuste sihtkatsete matkimine	15
1.4 Statistiline metoodika	17
2 Sihtkatse matkimine TÜ Eesti geenivaramu andmete põhjal	20
2.1 Ülevaade andmetest	20
2.1.1 Polügeenne riskiskoor	20
2.2 Sihtkatse matkimine	21
2.2.1 Protokolli komponendid	22
2.2.2 Valimi koostamine ja kirjeldav statistika	25
2.3 Andmeanalüüs	28
Kokkuvõte	32
Kasutatud allikad	34

Sissejuhatus

Käesoleva magistritöö eesmärk on selgitada sihtkatse matkimise teooriat, selle olulisust ja rakendamist. Töös kasutatakse Tartu Ülikooli Eesti geenivaramu andmeid sihtkatse protokollide koostamiseks ja rakendamiseks vaatlusandmetel, et hinnata kolesterooli alandavate ravimite ja polügeense riskiskoori mõju südame-veresoonkonna haiguste riskile.

Südame-veresoonkonna haiguste hulka kuulub mitmeid erinevaid diagnoose ning need kujutavad endast üht juhtivat surmapõhjust üle maailma. Aastal 2023 olid Eestis ligi pooled kõikidest surmapõhjustest tingitud vereringeelundite haigustest, kuigi suures osas on need ennetatavad õige ravi ja tervisliku eluviisiga. Südamehaiguste ja insuldi peamised käitumuslikud riskitegurid hõlmavad ebatervislikku toitumist, vähest kehalist aktiivsust ning tubaka ja alkoholi tarbimist. Nende ja teiste käitumuslike riskitegurite mõju võib avalduda kõrge vererõhu, vere glükoositaseme ja vere lipiidide taseme ning ülekaalu ja rasvumise vormis. (*Surmad. Statistikaamet 2024; Cardiovascular diseases. World Health Organisation 2021*) Seetõttu on oluline, et kõrge vererõhu tasemega inimesed saaksid vastavat kolesterooli alandavat ravi.

Töös kasutatakse vaatlusandmeid, et uurida, kuidas kolesterooli alandav ravi ja polügeenne riskiskoor mõjutavad kõrge kolesteroolitasemega inimeste seas kardiovaskulaarhaiguste esinemise riski. Uurimisküsimusele vastamiseks tuuakse välja teoreetiline taust sihtkatse matkimisest ja koostatakse analüüsi protokoll. Protokoll põhjal leitakse kogu kohordist sobiv valim ja teostatakse analüüs. Andmeanalüüsil kasutatakse elukestusanalüüsi meetodeid.

Töö esimeses peatükis keskendutakse randomiseeritud katse põhiaspektidele, selgitades selle peamisi teoreetilisi kaalutlusi. Samuti avatakse randomiseeritud sihtkatse matkimise teooriat ja tausta. Tutvustatakse ka lühidalt andmeanalüüsis kasutatud statistilist meetodikat. Teises peatükis viiakse läbi sihtkatse matkimise vajalikud sammud koos vastava analüüsiga Tartu Ülikooli Eesti geenivaramu andmete põhjal.

1 Sihtkatse matkimine vaatlusandmete põhjal

Ideaalsel juhul vastataks püstitatud uurimisküsimustele, nagu näiteks ravimi efektiivsus või turvalisus, kasutades sobivalt planeeritud ja läbi viidud randomiseeritud katset. Selliseid uuringuid peetakse sageli kõrgeima kvaliteediga tõendusmaterjaliks nende baastaseme segajate puudmise, analüüside lihtsuse ja põhjuslike mõjude osteste hinnagute tõttu (Kutcher *et al.*, 2021). Tegelikuses ei ole tihtipeale sellise uuringu teostamine näiteks eetiliste, rahaliste või ajaliste piirangute tõttu võimalik. Kasutades vaatlusandmeid, mis sageli kogutakse teistel eesmärkidel, saab neid piiranguid vähendada. Siiski ilmnevad ka selles takistused, sealhulgas asjaolu, et andmetes puudub randomiseeritud ravistrateegia määramine. Üldisemalt öeldes, kui randomiseeritud katse ja vaatlusandmetel põhinev uuring käsitlevad sarnast küsimust, võivad erinevused saadud hinnangute vahel olla seletatavad uuringu disaini ja teostatud andmeanalüüsi lahkevustega (Lodi *et al.*, 2019). Selliste, vaatlusandmetel põhinevate uuringute sagedane eesmärk ongi võimalikult hästi jäljendada hüpoteeetilist randomiseeritud katset. Seda hüpoteeetilist randomiseeritud katset, mida soovime matkida, nimetatakse ka sihtkatseks (ingl *target trial*).

Järgnevas jaos avame randomiseeritud katsete tausta, mida tuleb silmas pidada nende hüpoteeetiliste sihtkatsete planeerimisel. Peatükk põhineb Hernán ja Robins, 2010 raamatul ning Hernán ja Robins, 2016 artiklil, kui ei ole märgitud teisiti.

1.1 Randomiseeritud katse

Kuigi randomiseeritud katseid peetakse tõenduspõhiste seoste hindamise tipuks, on need siiski vastuvõtlikud erinevate nihete tekkimisele. Selles peatükis käsitletakse põhjuslike seoste olemust ja selgitatakse, miks randomiseerimine tagab usaldusväärsed põhjuslike seoste hinnangud. Samuti tuuakse välja randomiseeritud katsete olulisemad komponendid.

1.1.1 Põhjuslik seos

Olgu vaatluse all huvipakkuvad tunnused ekspositsioon X ja väljund Y . Meile pakub huvi, kas tunnusel X on põhjuslik mõju tunnusele Y (tähistatakse ka kui $X \rightarrow Y$). Tutvustame matemaatilist tähistust, mida kasutatakse põhjuslike seoste formaliseerimiseks.

Vaatame tunnust X kui ravi ehk ekspositsioon (1: ravisaja, 0: mitte-ravisaja) ja tunnust Y kui väljund-fenotüübi esinemine (1: esineb fenotüüp, 0: ei esine fenotüüpi). Võrreldes valimist hinnatavaid suurusid $P(Y = 1|X = 1)$ ja $P(Y = 1|X = 0)$, saame küll tuvastada seoseid tunnuste vahel, kuid mitte põhjuslikkust. Defineerime $Y^x := (Y|do(X) = x)$, on väljundi Y väärtus juhul, kui X oleks määratud olema kindla väärtusega x . Öeldakse, et ravil on individuaalne põhjuslik mõju, kui selle indiviidi puhul $Y^0 \neq Y^1$, ehk isiku väljund on erinev mõlema raviväärtuse korral. Binaarsete tunnuste puhul, esineb üldkogumis ravi ja fenotüübi vahel keskmine põhjuslik mõju (ingl *average causal effect*) kui

$$P(Y^1 = 1) \neq P(Y^0 = 1).$$

See tähendab, et samas populatsioonis kahe erineva raviväärtuse korral on erinev risk fenotüübi esinemiseks. Sellise mõju hindamiseks on vaja selgelt määratletud populatsiooni, kelle väljundeid Y^1 ja Y^0 võrreldakse. Pideva väljundtunnuse puhul esineb keskmine põhjuslik mõju, kui kehtib $E(Y^1) \neq E(Y^0)$. Üldkogumi keskmine põhjuslik mõju pideva väljundtunnuse puhul on defineeritud kui

$$E(Y^1) - E(Y^0).$$

Tasub täheldada, et keskmise põhjusliku mõju puudumine ei tähenda individuaalse mõju puudumist. See tähendab, et võib leida indiviidi, kelle väljund on erinev mõlema raviväärtuse korral. Üldkogumi keskmine põhjuslik mõju ongi keskmine individuaalne mõju $E(Y^1 - Y^0)$. Selliste väljundite võrdlemine indiviidide tasandil

ei ole enamasti küll võimalik, kuid üldkogumi keskmine põhjuslik mõju võib olla andmetest hinnatav. Samamoodi on hinnatavad näiteks ka üldkogumi keskmine põhjuslik mõju dispersioonile, mediaanile, riskile jne. Edaspidi viidatakse üldkogumi keskmisele põhjuslikule mõjule kui lihtsalt põhjuslik mõju.

1.1.2 Randomiseerimine

Põhjuslike järelduste tegemiseks on vaja andmeid sündmuste Y^1 ja Y^0 kohta, mis päris elus peaaegu kunagi korraga ei toimu (seetõttu nimetatakse neid ka kontrafaktuaalseteks tunnusteks). Küsimus seisnebki selles, kuidas on võimalik kasutada päriselulisi andmeid põhjuslike järelduste tegemiseks. Lahenduseks oleks randomiseeritud katse teostamine. Selles peatükis vaatame, kuidas võimaldab randomiseerimine tuvastada põhjuslikke seoseid.

Viies läbi randomiseeritud katse saame siiski andmed, kus iga indiviidi kohta on teada vaid ühe kontrafaktuaalse väljundi väärtus. Randomiseerimise abil tagame, et see millisse ekspositsiooni X rühma individ kuulub on juhuslik. See tähendab ka, et ekspositsioon ei ole seotud segavate tunnustega. Näiteks võib mõelda, et oleme üldkogumis juhuslikult määranud kõik inimesed kahte gruppi. Nüüd võime vabalt valida, kummale grupile ravi määrata ja $P(Y = 1|X = 1)$ ja $P(Y = 1|X = 0)$ väärtused sellest ei muutu. See tähendab, et risk potentsiaalse ravi väärtuse x korral esimeses grupis, $P(Y^x = 1|X = 1)$, on võrdne riskiga potentsiaalse ravi väärtuse x korral teises grupis, $P(Y^x = 1|X = 0)$, mõlema $x = 0$ ja $x = 1$ korral. Seda tingimust nimetatakse ka vahetatavuseks (ingl *exchangeability*) ja tähistatakse kui $Y^x \perp\!\!\!\perp X$. Kokkuvõttes tähendab see, et ravisaajad ja mitte-saajad kogeksid sama riski, kui neile oleks määratud sama ekspositsiooni tase ($x = 0$ või $x = 1$). Seega marginaalne risk ravi väärtuse x korral on

$$P(Y^x = 1|X = 1) = P(Y^x = 1|X = 0) = P(Y^x = 1).$$

Järelikult saame valimist hinnata $P(Y^1 = 1)$, kuna see on võrdne väärtusega

$P(Y = 1|X = 1)$ ning sama kehtib ka $P(Y^0 = 1)$ kohta. Sellisest ideaalsest randomiseeritud katsest tuvastatud seos ongi põhjuslik.

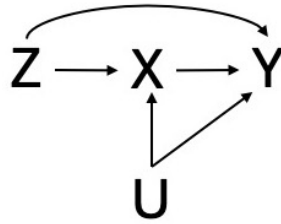
1.1.3 Randomiseeritud katse komponendid

Enne randomiseeritud katse või selle jäljendamise läbi viimist on vaja paika panna uuringu komponendid. Nendeks on valimisse sobivuse kriteeriumid, ravistrateegiad, ravistrateegiate määramine, jälgimisaeg, huvipakkuvad sündmused, huvipakkuvad põhjuslikud mõjud ja analüüsi plaan. Esiteks vaatame hüpoteetilise uuringu ehk sihtkatse protokollid ja hiljem keskendume sellele, kuidas seda uuringut jäljendada vaatlusandmete põhjal.

Valimisse sobivuse kriteeriumite (ingl *eligibility criteria*) alla võivad kuuluda näiteks demograafilised andmed, haiguste olemasolu või diagnostiliste testide tulemused. Kõik valimisse sobivuse kriteeriumid peavad olema täidetud enne jälgimisaja algust, et kaasata indiviid uuringusse.

Ravistrateegiad on määratletud kui sekkumised, mille mõju meid huvitab. Näiteks võivad ravistrateegiad hõlmata ravimi ($X = 1$) või platseebo ($X = 0$) saamist. Peale individide tuvastamist, kellel on kõik valimisse sobivuse kriteeriumid täidetud, **määratakse neile ravistrateegia**. Randomiseeritud katses toimub see määramine juhuslikult, tagades sellega vahetatavuse, mis võimaldab hinnata andmetest põhjuslikku mõju. Kindlustamiseks, et ainus erinevus randomiseeritud gruppide vahel seisneks määratud ravis, kasutatakse vahel ka pimesi ravi määramist (ingl *blind assignment*). See tähendab, et patsient ega raviarst ei ole teadlikud sellest, kas määratud ravim on aktiivne või platseebo.

Jälgimisaeg algab nullpunktist, mis on hetk, mil valimisse sobivuse kriteeriumid on täidetud ja toimub ravistrateegia määramine (randomiseerimine). Sellest punktist algab uuringu huvipakkuvate sündmuste jälgimine. Jälgimisaeg lõppeb huvipakkuvatest sündmustest, tsenseerimisest või uuringu lõpust esimese esinemisel.



Joonis 1: Graafiline estilis uuringust, kus individidele määratakse ravi Z , manustatakse ravi X ja uuritakse väljundina sündmust Y . U kujutab mõõtmata riskifaktoreid, mis mõjutavad tunnust X ja Y .

Kõik randomiseeritud katsed hindavad väljundeid ehk **huvipakkuvad sündmused**. Enamikul uuringutel on mitu tulemust, millest mõned on suurema huvi all kui teised. Kõik huvipakkuvad sündmused tuleb enne uuringu algust ja patsientide kaasamist selgelt määratleda. (Kutcher *et al.*, 2021)

Randomiseeritud katsel võib olla mitu **huvipakkuvat põhjuslikku seost** (ingl *causal contrast of interest*). Kaks sagedast sellist seost on ravi kavatsuse mõju (ingl *intention-to-treat effect*), ehk ravistrateegia määramise mõju sõltumata sellest, kas isik jätkab selle järgimist. Teine on protokoll järgimise mõju (ingl *per-protocol effect*), ehk ravistrateegia mõju kui järgitakse uuringuprotokollis määratud raviplaani. Sageli pakuvad huvi mõlemad mõjud.

Nende mõjude paremaks mõistmiseks vaatame olukorda, nagu on visualiseeritud joonisel 1. Olgu Z uuringus osalevale individile määratud raviplaan (1: kuulub ravigruppi, 0: ei kuulu ravigruppi). Tunnus X kujutab aga tegelikult saadud ravi (1: sai ravi, 0: ei saanud ravi). Isikule määratud ravi võib erineda tegelikult saadud ravist mitmel põhjusel. Need on kajastatud tunnuses U mõõtmata tunnustena. Samad tunnused võivad mõjutada ka huvipakkuva sündmuse esinemist ja on siis ka mõõtmata riskifaktorid. Samuti võib isikule määratud raviplaan mõjutada huvipakkuva sündmuse esinemist ka ilma X vahenduseta (seos $Z \rightarrow Y$). Näiteks võib teadlikkus määratud ravist muuta indiviidi käitumismustreid. Ravi kavatsuse mõju, mida sageli randomiseeritud katsetes käsitletakse, ongi raviplaani Z mõju väljundile Y .

Sellised ravi kavatsusel põhinevad hinnangud sõltuvad randomiseeritud raviplaani kinnipidamise määrast.

Protokolli järgimise mõju puhul vaatame olukorda, kus uuringusse kaasatud in-diviivid järgivad neile määratud ravi täielikult. Sellisel juhul huvitab meid seos $X \rightarrow Y$. Kui seose $X \rightarrow Y$ uurimisel eksisteerib selline tunnus U , nagu toodud joonisel 1, tuleb protokolli järgimise mõju uurimisel kohandada sellele tunnusele. Kui see tunnus on mõõtmata, ei ole põhjuslikku mõju võimalik hinnata lihtsatel meetoditel ega ilma täiendavate eeldusteta.

1.2 Vaatlusuuringud ja kehtivate põhjuslike järelduste tingimused

Ideaalsed randomiseeritud katsed võimaldavad põhjuslike seoste otsest hindamist, kuid vaatlusandmete puhul ei kehti samad tingimused nagu randomiseeritud katsetes. Üks nendest tingimustest on juba tutvustatud vahetatavus. Ilma randomiseeritud ravistrateegia määramiseta võib leida tunnuseid, mis suurendavad ravi määramise tõenäosust ja ka väljundi ilmnemise riski. See tähendabki, et ravisajad ei oleks kogunud sama riski kui mittesajad, kui poleks kellegile ravi määratud. Parim lähenemine sellises olukorras on siiski teostada analüüs eeldusel, et ravi X saadakse juhuslikult, kontrollides üle mõõdetud tunnuste L , mis mõjutavad ravi ja väljundit. Selliseid tunnuseid L nimetatakse ka segajateks (ingl *confounders*). Siis öeldakse, et kehtib tinglik vahetatavus (ingl *conditional exchangeability*) $Y^x \perp\!\!\!\perp X|L$ ja sellist uuringut on võimalik vaatlusandmetest jäljendada järgmistel tingimustel:

1. Võrreldavad ravistrateegiad vastavad hästi defineeritud sekkumistele, mis omakorda vastavad andmetes kajastatud ravile. Tingimust nimetatakse järjepidevuseks (ingl *consistency*).
2. Tinglik tõenäosus iga ravi X väärtuse esinemiseks sõltub ainult mõõdetud

tunnustest L . Sellega tagatakse eelpool toodud tingliku vahetatavuse tingimus.

3. Iga ravi X taseme esinemise tõenäosus tingimusel L on suurem nullist. Tingimusele viidatakse ka kui positiivsusele (ingl *positivity*).

Kui mõni eelpool toodud tingimustest ei kehti, ei saa vaatlusandmetest jäljendada tinglikult randomiseeritud katset.

1.3 Sihtkatse matkimine

Edukalt matkitud sihtkatse võimaldab saada vaatlusandmetest samu hinnanguid, mis oleks saadud randomiseeritud katsest. Selle saavutamiseks on vajalik hoolikas uuringu disaini ja analüüsi planeerimine. Erinevused saadud hinnangute vahel võivad tuleneda näiteks kontrollimata jäetud segajatest või lahknevusteset uuringu protokollides. Järgnevas peatükis kirjeldatakse sihtkatse matkimise (ingl *target trial emulation*) ülesehitust, tutvustatakse võimalikke nihkeallikaid ja ühte võimalikku erivarianti sihtkatse matkimisest.

1.3.1 Ülesehitus

Meenutame peatükis 1.1.3 käsitletud randomiseeritud katse komponente ning uurime põhilisi erinevusi, mis ilmnevad nende matkimisel. Peamine, millest alustada, on korrektse põhjusliku küsimuse sõnastamine ja soovitud sihtkatse paika panemine. Oluline on, et vaatlusandmestik sisaldaks piisavalt teavet, et kõiki protokollide komponente korrektselt rakendada. Vaatame järgnevalt uuringu protokollide komponentide matkimist täpsemalt.

Vaatlusandmetel põhinev uuring peaks rakendama samu **valimisse kaasamise kriteeriumeid**, nagu sihtkatse. Nende kriteeriumite rakendamiseks on vaja andmestikus ajaloolisi andmeid. Protokollis ei tohiks valimisse kaasamise kriteeriumite

all olla sündmusi, mis toimuvad peale jälgimisaja algust (näiteks „valimisse kaasata ainult inimesed, kes on jälgimisaja jooksul on saanud ravi“). Lisaks ei ole võimalik isikuid peale jälgimisaja algust enam valimist välja jätta. Selline olukord tekib näiteks siis, kui isik kolib riigist ära ja tema kohta puuduvad andmed sellest hetkest alates. See isik tuleb pidada siis tsenseerituks, mitte valimist välja jäetunuks.

Vaatlusandmetel põhineval uuringul on piirangud selle osas, milliseid **ravistrateegiaid** saab võrrelda. See tähendab, et võrrelda saab ainult selliseid strateegiaid, mis on tegelikult kasutusel. Näiteks ei ole võimalik võrrelda strateegiaid alustas ravi ($X = 1$) ja alustas platseebot ($X = 0$). Samuti on vaatlusandmetes harv nähtus, et kohe jälgimisaja alguses toimub **ravistrateegia määramine**, nagu see toimuks randomiseeritud katses. Seetõttu võidakse matkitud sihtkatses panna paika armuaeg (ingl *grace period*), mille jooksul peale jälgimisaja algust võib isik ravi alustada ja siiski kuuluda ravisaajate hulka. Näiteks võib ravistrateegia olla defineeritud kui „alustada ravi 3 kuu jooksul peale jälgimisaja algust.“ Armuaaja defineerimine tagab selle, et ravistrateegiad jäävad realistlikuks ning suurendab ka nendne inimeste arvu, keda saame sihtuuringu jäljendamiseks kaasata ravisaajate hulka.

Ravistrateegia määramise puhul on oluline ka see, et me saame matkida uurin-guid ainult ilma pimesi ravi määramiseta. Päriselulises stsenaariumis ei ole võimalik, et valimi subjektid ja tervishoiutöötajad ei oleks teadlikud patsiendile määratud ravist. Selleks, et matkida juhuslikku ravistrateegiate määramist, on vajalik kohandamine segajatele, nagu on täpsustatud peatükis 1.2. Eesmärk on tagada tinglik vahetatavus eelnevalt defineeritud ravistrateegia gruppides. Kui andmetes ei ole piisavalt informatsiooni nende segajate kohta, või kui me ei suuda neid korrektselt paika panna, ei ole sihtkatse matkimine võimalik.

Jälgimisaja alguse, ehk nullpunkti paika panemine on edukalt matkitud sihtkatse aluseks. Valimisse sobivuse kriteeriumid peavad olema täidetud selleks hetkeks, kuid mitte hiljem. Uuringu huvipakkuvaid sündmusi hakatakse jälgima sellest aja-

hetkest, kuid mitte varem. Üldiselt peab jälgimisaeg vaatlusuuringutes algama samal ajal kui randomiseeritud katses. Selle tingumuse rikkumine võib aga kergesti juhtuda ja oluline on seega täielik arusaamine matkitavast protokollist. Järgnevas peatükis välja toodud nihked on samuti tekitatavad nullpunkti valesti defineerimisega.

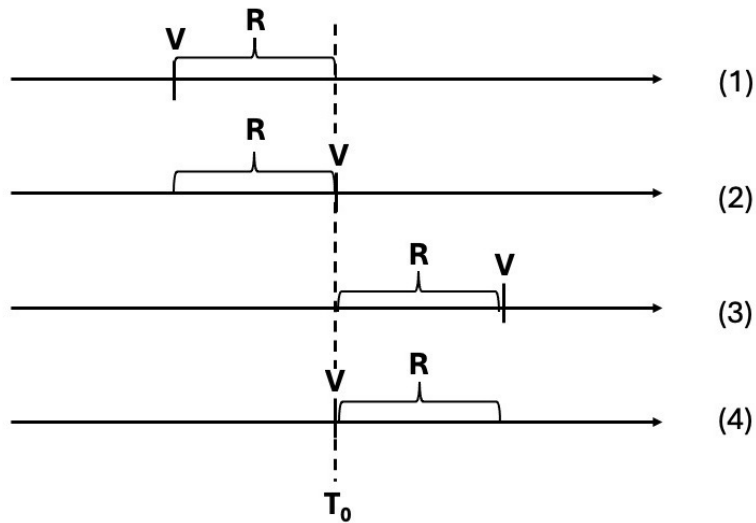
Sihtkatse matkimisel on nagu randomiseeritud katseski oluline, et **huvipakkuvad sündmused** oleksid selgelt defineeritud sihtkatse planeerimise etapis. Kõige eelistatum oleks sellise sihtkatse matkimine, kus toimub huvipakkuvate sündmuste süstemaatiline ja pime hindamine. Sellega on tagatud, et teadmised patsiendi ravistaatusest ei mõjuta arsti otsust otsida sündmust. Näiteks võib raviarst teostada sagedasemaid rinnavähikontrolle hormoonravi saajate seas, mis võib viia suurenenud diagnooside arvuni. Siiski, kuna patsiendid ja tervishoiudtöötajad on tavaliselt teadlikud saadud ravist, ei saa me kasutada vaatlusandmeid süstemaatilise ja pimesi sündmuste hindamisega sihtkatse matkimiseks. Erandiks on juhud, kus sündmuste hindamist ei mõjuta raviajalugu, näiteks kui tulemuseks on surm.

1.3.2 Võimalikud nihkeallikad

Ideaalses randomiseeritud katses on iga indiviidi nullpunkt see hetk, mil toimub randomiseerimine ja valimisse sobivuse kriteeriumid on täidetud. Vaatlusuuringute korral on oluline nullpunkti, valimisse sobivuse kriteeriumite ja ravistrateegia määramise sünkroniseerimine nii, nagu seda oleks tehtud sihtkatses. Selle põhimõtte rikkumine on üks peamisi põhjuseid, miks saadakse ka nihkega hinnangud. Joonisel 2 on toodud erinevad vead, mis võivad tekkida sihtkatse matkimisel.

Joonise 2 teljel (1) on kujutatud olukord, kus jälgimisaeg algab peale valimisse sobivuse kriteeriumi täitmist ja ravistrateegia määramise perioodi. Sellisest analüüsist saadud hinnangud on nihkega, kuna analüüs piirdub nende inimestega, kes jäid uuringusse kuni jälgimisaja alguseni (ingl *prevalent user bias*).

Teine sihtkatse matkimisel tehtav viga, mis on kujutatud joonise 2 teljel (2), on olu-



Joonis 2: Võimalikud vead nullpunkti defineerimisel sihtkatse matkimisel. V tähistab valimisse sobivuse kriteeriumi täitmist, R ravistrateegia määramise perioodi ja T_0 nullpunkti.

kord, kus valimisse sobivus ja jälgimisaja algus toimuvad koos, kuid ravistrateegia määratakse varem. Sellisel juhul, kui isik ei vastanud valimisse sobivuse kriteeriumitele ravi määramisel ei saa me olla kindlad, et ravi ei mõjutanud neid kriteeriume hiljem vastavaks (ingl *post treatment selection bias*). (Kutcher *et al.*, 2021)

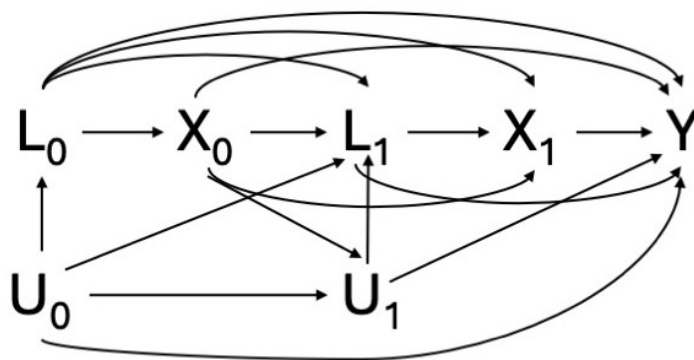
Joonise 2 teljel (3) on sama oht valikunihke tekkimiseks, nagu eelmisel juhul, kuna valimisse sobivuse kriteeriumid peavad olema täidetud peale ravi määramist. Lisaks on tekkinud olukord, kus ravisaajatel ei saa jälgimisaja mingi perioodi jooksul sündmust esineda. Sellist perioodi nimetatakse ka surematuks ajaks (ingl *immortal time*), mille jooksul on risk sündmuse esinemiseks ravisaajaate seas on garanteeritud null. (Hernán *et al.*, 2016)

Viimasel juhul joonisel 2 on kujutatud klassikaline surematu aja nihe (ingl *classical immortal-time bias*), kus on korrektselt defineeritud küll nullpunkt ja valimisse sobivuse kriteeriumid, kuid ravistrateegia määramine toimub tingimustel, mis juhtuvad peale nullpunkti. Siis on eelmise olukorraga analoogselt ravisaajatel uuringus surematu aeg, mil sündmust toimuda ei saa. (Hernán *et al.*, 2016) Mida pikem on

uuringus lubatud surematu aeg, seda suurem nihe tekitatakse.

1.3.3 Järjestikuste sihtkatsete matkimine

Eelnevalt rääkisime ravistrateegia määramise ja nullpunkti ühildamise olulisusest ning ka armuaja defineerimisest olukordades, kus ravi määramine ei saa toimuda täpselt jälgimisaja alguses. Armuaja määramine on üks viis, kuidas vältida surematu aja nihke teket, kuid sellega kaotame andmetes olevat informatsiooni. Näiteks, kui oleme ühe ravistrateegiana defineerinud „alustada ravi 3 kuu jooksul peale jälgimisaja algust,“ kuuluvad alternatiivsesse ravistrateegiasse ka kõik need isikud, kes alustasid ravi hiljem kui 3 kuud peale jälgimisaja algust. Üks viis, kuidas siiski kõiki saadavalolevaid andmeid ravisaajate kohta rakendada, on järjestikuste sihtkatsete matkimine (ingl *sequential target trial emulation*). Joonisel 3 on toodud näide ühest järjestikuseid sihtkatseid illustreerivast graafist, mida soovime vaatlusandmete põhjal matkida.



Joonis 3: Kahe vaatlusajaga, $k = 0$ ja $k = 1$, järjestikused sihtkatsed. Toodud on mõõdetud L_k , mõõtmata U_k ja ekspositsiooni X_k tunnused vaatlusaegadel k . Y on huvipakkuv väljund.

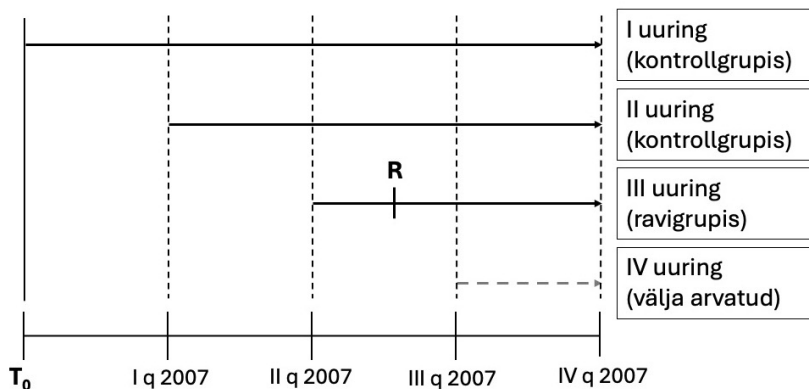
Vaatleme ajas muutuvat binaarset tunnust X_k , mis vastab ravi staatusele ajal k . Esimene võimalik hetk ravi alustamiseks on ajal $k = 0$ ja viimane ajal $k = K$, kus $K + 1$ on järjestikuste katsete arv. Samamoodi tähistab L_k mõõdetud ja U_k

mõõtmata tunnuseid ning Y on huvipakkuv väljund. Joonisel 3 on lihtsuse mõttes visualiseeritud ainult kaks katset ehk kaks jälgimisaega $k = 0$ ja $k = 1$. Sellised järjestikused sihtkatsed on vaatlusandmetest jäljendatavad, kui mõõtmata tunnused U_k pole segajad (vaata peatükk 1.2). Ravi määramine ajahetkel k sõltub eelmisest ravi staatusest X_{k-1} ja mõõdetud segajatest L_0, \dots, L_k . Kusjuures, individid ei ole alustanud ravi enne uuringu algust, ehk $X_{-1} = 0$. Järjestikustes sihtkatsetes võib individid panustada igasse katsesse alustades ajahetkest $k = 0, \dots, K$, mil valimisse sobimise kriteeriumid on täidetud.

Varem rääkisime, kuidas saame matkida sihtkatsed, kus kehtib tinglik vahetatavus $Y^x \perp\!\!\!\perp X|L$. Ajas muutuva ravistaatusega on samuti vajalik kontrollimine tunnustele L_k , et saavutada tinglik vahetatavus igal ajahetkel ehk järjestikune tinglik vahetatavus (ingl *sequential conditional exchangeability*). Eelmises näites, kus on kaks vaadeldavat ajahetke, on järjestikune tinglik vahetatavus kombinatsioon tinglikust vahetatavusest mõlemal ajahetkel $k = 0$ ja $k = 1$. Peale vahetatavuse on järjestikuste sihtkatsete matkimisel olulised ka järjestikused versioonid positiivsuse ja järjepidevuse tingimustest.

Joonisel 4 on toodud lihtsustatud skeem, kuidas käib järjestikuste sihtkatsete matkimine vaatlusandmetel. Selle meetodi rakendamiseks on oluline, et individidel oleks võimalik vastata valimisse sobivuse kriteeriumitele rohkem kui korra jälgimisaja jooksul. Indiviid võib panustada uuringusse mitmel korral, kui ta vastab valimisse sobivuse kriteeriumitele ka peale jälgimisaja algust (näiteks ühel või mitmel joonisel 4 punktiiriga tähistatud punktis). Indiviidid, kes on enne k -ndat hetke saanud ravi, ei vasta enam valimisse sobivuse kriteeriumitele uuringu jaoks, mis algab hetkest k . Selline olukord on illustreeritud joonisel 4 uuringus 4, kus indiviid ei osale enam uuringus, kuna talle määrati ravi eelmises uuringus. Seega, isikud kes on kaasatud uuringusse algusega k sisaldavad neid, kes

- (a) alustavad ravi ajavahemikus k kuni $k + 1$;
- (b) jäävad ajahetkeks $k + 1$ endiselt ravi mitte-saajateks.



Joonis 4: Järjestikuste randomiseeritud katsete matkimise skeem indiviidil, kelle jälgimisaeg algab jaanuaris 2007, mis on ka I uuringu algus. Ravistrateegia on „alustada ravi kuni 3 kuud peale jälgimisaja algust.“ R tähistab ravi määramist, T_0 nullpunkti ja nool indiviidi jälgimisaega uuringus.

Igas uuringus võrreldakse neid kahte kohorti. Ravi mitte-saajad uuringus peale ajahetke k võivad esineda ravisaajatena mõnes järgnevas uuringus algusega $k + 1, k + 2, \dots, K$. Indiviid saab ravi saajana esineda ainult ühes uuringus, kuid mitte ravi-saajana mitmes. (Keogh *et al.*, 2023)

1.4 Statistiline metoodika

Üks sagedaseim põhjuslik küsimus, millele vastust otsitakse, puudutab ravi mõju sündmuse toimumiseni kulunud ajale. Selliste andmete analüüsimiseks, kus huvipakkuv sündmus on mingist alghetkest kuni sündmuse esinemiseni, kasutatakse elukestusanalüüsi. Neid andmeid nimetatakse ka elukestusandmeteks ja need võivad olla tsenseeritud juhul, kui indiviidil pole jälgimisaja jooksul sündmust esinenud.

Elukestusanalüüs on rakendatav ka sihtkatsete matkimise kontekstis ja see tugineb elukestuse jaotusele. Selle iseloomustamiseks on ühe põhilise funktsioonina defineeritud riskifunktsioon. Riskifunktsioon $h(t)$ iseloomustab riski, et sündmus toimub teatud ajahetkel t . See funktsioon saadakse tõenäosusest, et indiviidil esineb sündmus ajahetkel t tingimusel, et see ei toimunud enne seda momenti. Täpsemalt on

riskifunktsiooni kuju defineeritud järgmiselt:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t},$$

kus juhuslik suurus T kirjeldab indiviidi elukestust. Elukestusanalüüsi läbiviimise üks peamine eesmärk on tuvastada, millised tunnused ja nende kombinatsioonid mõjutavad riskifunktsiooni. Ühe võimalusena uurime Coxi võrdeliste riskide mudelit. Mudel eeldab, et riskifunktsioon i -nda indiviidi jaoks on esitatav kui

$$h_i(t) = \exp(\boldsymbol{\beta}' \mathbf{x}_i) h_0(t),$$

kus $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$ on parameetrite vektor p seletavale tunnusele, mille väärtused i -ndal indiviidil on $\mathbf{x}_i = (x_{1i}, x_{2i}, \dots, x_{pi})$. Funktsioon $h_0(t)$ on riskifunktsioon indiviidile, kelle puhul on kõigi seletavate tunnuste väärtus võrdne nulliga. Coxi võrdeliste riskide mudelis funktsiooni $h_0(t)$ kuju ei määrata ja hinnatakse vaid parameetrid $\boldsymbol{\beta}$. (Collett, 2023)

Coxi mudel on hinnatav ka järejestikuste sihtkatsete kontekstis. Selleks on andmed vaja tavapärase laia formaadi asemel, kus iga isik vastab ühele reale, teisendada kujule, kus iga rida andmestikus vastab ühe indiviidi ühele uuringule. Sellise andmestiku iga rida vastab ühele isiku-uuringule (ingl *person-time*). Nii on võimalik andmetest hinnata Coxi mudelist riskisuhteid. Kuna aga üksikisikud võivad olla kaasatud mitmesse jäljendatud uuringusse, on vajalik tavapärase dispersiooni hinnangu vastav kohandamine. Üks viis selleks on kasutada *bootstrap*-meetodit. Parameetrite hinnangutele vastavate p -väärtuste leidmiseks arvutatakse uuele dispersiooni hinnangule kohandatud Waldi statistik. Nullhüpoteesi korral eeldatakse, et parameeter ei erine statistiliselt oluliselt nullist ning statistik järgib siis normaaljaotust.

Bootstrap-meetod põhineb eeldusel, et olemasolev valim kirjeldab üldkogumit. Vaatame täpsemalt, kuidas leida *bootstrap*-hinnang Coxi mudelist leitud parameetri

β_i , $i = 1, 2, \dots, p$ hinnangu standardveale kasutades Monte Carlo algoritmi. Ütleme, et oleme leidnud mudelist hinnangu $\hat{\beta}_i$. Monte Carlo algoritmi rakendamine hõlmab kolme sammu:

1. valimist mingi arvu B tagasipanekuga juhuslike alamvalimite võtmine. Neid alamvalimeid nimetatakse ka *bootstrap*-valimiteks. *Bootstrap*-valimite suurus on esialgse valimi maht.
2. Iga *bootstrap*-valimi puhul huvipakkuva hinnangu $\hat{\beta}_i^b$, $b = 1, 2, \dots, B$ leidmine osalise tõepära funktsiooni maksimeerimise kaudu.
3. $\hat{\beta}_i^b$, $b = 1, 2, \dots, B$ standardhälbe leidmine vastavalt valemitele

$$\hat{\sigma}_B = \sqrt{\frac{\sum_{b=1}^B (\hat{\beta}_i^b - \hat{\beta}_i^{*})^2}{B-1}}, \quad \hat{\beta}_i^{*} = \frac{\sum_{b=1}^B \hat{\beta}_i^b}{B}.$$

Kui $B \rightarrow \infty$, läheneb $\hat{\sigma}_B$ *bootstrap*-hinnangule $\hat{\beta}_i$ standardveast, mis on suurima tõepära hinnang tegelikule standardveale. (Efron ja Tibshirani, 1985)

2 Sihtkatse matkimine TÜ Eesti geenivaramu andmete põhjal

2.1 Ülevaade andmetest

Tartu Ülikooli Eesti geenivaramu (TÜ EGV) on riiklik biopank, millega on 2023. aasta seisuga liitunud üle 210 000 inimese. Geenidoonorite andmeid on kogutud juba alates 2002. aastast ja see kaasab ligikaudu 20% Eesti täiseelisest elanikkonnast (*Üldinfo. Tartu Ülikooli Eesti geenivaramu 2023*). Geenivaramu doonorite koosseis kajastab üldiselt eesti elanikkonna soolist, vanuselist ja geograafilist jaotust (*Eesti geenivaramu. Tartu Ülikool 2021*).

Peale pärilikkuseaine (DNA) sisaldab biopank andmeid ka doonorite tervisliku seisundi, keskkondlike tegurite, elustiili ning toitumise kohta. Geenidoonoritest ligikaudu 65,4% on naised ja mõni vereringeelundite haigus (kõrgvererõhkhaigus, südame isheemiatõbi, muu südamehaigus ja/või peaaajuveresoonte haigus) on diagnoositud ligikaudu 87 000 doonoril. Andmebaasi täiendatakse perioodilistelt Eesti Tervisekassa andmetest haiguste olemasolu ja diagnooside kuupäevadega. Tervise Arengu Instituudi andmete põhjal uuendatakse surma põhjuseid ja -kuupäevi. Viimane uuendus mõlemast andmebaasist antud analüüsil toimus 2022. aasta lõpus.

2.1.1 Polügeenne riskiskoor

Geenidoonoritele on leitud mitmeid polügeenseid riskiskoores (PRS), mis aitavad hinnata indiviidide geneetilist riski huvipakkuvale fenotüübile (näiteks haiguse esinemisele). Nende seas on ka riskiskoor südame-veresoonkonna haigustele, mis meile antud uuringus huvi pakub.

Polügeensete riskiskooride väljatöötamine tugineb ühenukleotiidsetel polümorfismidel (SNPde) põhineva geneetilise varieeruvuse mõju kvantifitseerimisel. Kuna üksikutel SNPdel on väike mõju geneetilisele varieeruvusele, kasutatakse polügeense

riskiskoori leidmiseks erinevate SNPide alleelidooside summat. Alleelidoos on genotüübis valitud alleeli esinemiste arv ja genotüüp on kõigi organismi kromosoomides paiknevate geenialleelide kogum. Eeldatakse, et iga SNP jaoks on kolm võimalikku genotüüpi. Olgu näiteks genotüübid aa, Aa ja AA. Kui valime lugemiseks alleeli A, on alleelidoosid siis vastavalt alleeli A esinemise arvud 0, 1 ja 2. Üldiselt on polügeenne riskiskoor i -nda indiviidi jaoks defineeritud kui k SNPi kaalutud summa:

$$GRS_i = \sum_{j=1}^k \omega_j X_{ij},$$

kus X_{ij} on i -nda indiviidi j -nda SNPi alleelidooside arv ja $\omega_j \in (-\infty, \infty)$ on j -nda SNPi kaal. (Läll, 2019) Käesolevas töös kasutatud PRS on leitud Patel *et al.*, 2023 artiklis avaldatud metaanalüüsi tulemuste põhjal polügeensete riskiskooride kataloogi *PGS Catalog* abil.

2.2 Sihtkatse matkimine

Uuritakse kolme huvipakkuvat sündmust eraldi, et leida kolesterooli alandava ravi ja polügeense riskiskoori mõju südame-veresoonkonna haiguste riskile:

1. Aeg esimese esinemiseni sündmustest: diagnoos I20-I25¹ (südame-isheemiatõved), I63 (peaaajuinfarkt) välja arvatud I63.6 (peaaajuinfarkt peaaaju mittepüogeense venoosse tromboosi tõttu) või surm südame-veresoonkonna haigusesse;
2. Aeg esimese esinemiseni sündmustest: diagnoos I21, I22 (äge või korduv müokardiinfarkt), I63 (v.a I63.6) või surm südame-veresoonkonna haigusesse;
3. Aeg surmani südame-veresoonkonna haigusesse.

¹RHK-10 (rahvusvahelise haiguste klassifikatsiooni) koodid, mille vasted on leitavad [Sotsiaalministeerium 2023](#) lehelt.

Edaspidi viidatakse eelnevatele sündmustele vastavalt kui esimene, teine ja kolmas huvipakkuv sündmus või kokkuvõtvalt kardiovaskulaarsündmused.

Uuringus vaadeldavad lipiidide sisaldust vähendavad ravimid kuuluvad statiinide klassi. Selleks, et hinnata polügeense riskiskoori ja kolesterooli alandava ravi mõju südame-veresoonkonna haiguste riskile, kasutame TÛ EGV vaatlusandmeid sihtkatse matkimiseks. Antud uuringus on huvipakkuvaks põhjuslikuks seoseks ravi kavatsuse mõju südame-veresoonkonna haiguste riskile kõrge kolesterooliga isikute seas. Samuti uuritakse polügeense riskiskoori mõju, kohandades seda ravistrateegiate mõjule. Kuna kohordis olevatel inimestel on võimalik valimisse kaasamise kriteeriumitele vastata mitmel korral jälgimisaja jooksul, teostame järjestikuste randomiseeritud sihtkatste matkimise. Nii tagame maksimaalse ravisaajate info kaasamise ja suurendame hinnangute täpsust.

2.2.1 Protokollid komponendid

Esimene samm sihtkatse matkimisel on täpsustada uuringu protokoll, mis olemasolevate vaatlusandmete raames oleks ideaaljuhul läbi viidud. Polügeense riskiskoori ja kolesterooli alandava ravi mõju hindamiseks südame-veresoonkonna haiguste riskile, on tabelis 1 välja toodud TÛ EGV andmete põhjal koostatud protokollid kokkuvõtte.

Protokollis on toodud kuus põhilist sihtkatse protokollid ja nende kirjeldused sihtkatse matkimisel TÛ EGV andmetel. Protokollid on koostatud vastavuses sihtkatse matkimise põhimõttega, et jälgendatud uuringu protokollid komponendid oleks võimalikult lähedases vastavuses komponentidga, mis kehtiksid sihtkatse läbi viimisel. Protokollid koostamisel on silmas peetud peatükis 1.3.2 kirjeldatud võimalikke nihkeallikaid ning protokollid on koostatud järjestikuste sihtkatsete matkimise jaoks.

Tabel 1: Sihtkatse matkimise protokollide kokkuvõtte hindamiseks polügeense riskiskoori ja kolesterooli alandava ravi mõju südame-veresoonkonna haiguste riskile.

Protokolli komponent	Sihtkatse protokollide kirjeldus	Jälgendus vaatlusandmetel kasutades TÕ EGV andmeid
Valimisse sobivus	<ul style="list-style-type: none"> • Isiku kolesterooli tase on mõõdetud ajavahemikus 2010-2019. • Isiku kolesterooli tase on piisavalt kõrge kolesterooli alandava ravi saamiseks (LDL-kolesterool ≥ 4 mmol/L või üldkolesterool $\geq 6,5$ mmol/L). • Enne esimest kõrge tasemega kolesterooli mõõtmist pole välja kirjutatud ühtegi kolesterooli alandava ravi retsepti. • Pole diagnoositud südame-veresoonkonna haigusi enne esimest kõrge tasemega kolesterooli mõõtmist või enne 2010 aastat. 	<p>Sihtkatseid matkitakse TÕ EGV andmetel igal järjestikusel aastal vahemikus 2010-2019. Igal aastal rakendatakse doonoritele samu valimisse kaasamise kriteeriumeid, mis kehtiksid sihtkatseski. Kui doonoril puudub kolesterooli mõõtmine mõne järjestikuse katse aastal, leitakse väärtus(ed) interpoleerimise teel. Vähendamaks andmesisestusvigade mõju, ei tohi LDL-kolesterooli tase olla kõrgem kui samal päeval mõõdetud üldkolesterool ja nende erinevus ei tohi olla suurem kui 4 mmol/L.</p>
Ravistrateegiad	<ol style="list-style-type: none"> 1. Ravisaajate hulka kuuluvad isikud, kes on kuni aasta peale jälgimisaja algust saanud kolesterooli alandava ravi retsepti. 2. Mitte-ravisaajate hulka kuuluvad isikud, kellele ei ole välja kirjutatud kolesterooli alandava ravi retsepti või on saanud retsepti rohkem kui aasta peale jälgimisaja algust. 	<p>Jälgendatud järjestikuste uuringute ravistrateegiad on samad, mis sihtkatseski, kuid vaatlusandmetes ei pruugi doonorid retsepti välja osta, ega raviplaanist täielikult kinni pidada.</p>

Ravistrateegia määramine	Patsiendid on juhuslikult määratud ravistrateegiateesse. Patsient ja teda raviv arst on teadlikud määratud ravistrateegiast.	Kolesterooli alandava ravi määrab doonorit raviv arst ja see on kajastatud vaatlusandmetes. Patsient on teadlik talle määratud ravistrateegiast. Ravistrateegia määramine on eeldatud, et toimub juhuslikult üle iga uuringu baas-taseme tunnuste: sugu, haridus, vanus, suitsetamine, LDL-kolesterool, diabeet, KMI, kõrgvererõhktõbi, lipoproteiini-ainevahetuse häire, krooniline neerupuudulikkus, kalendriaasta.
Jälgimisaeg	Iga valimisse sobiva isiku jälgimisaeg algab randomiseerimisest ja lõpeb järgnevatest sündmustest esimese toimumisel: surm, tsenseerimine või huvipakkuva sündmuse esinemine.	Sama, kuid jälgimisaeg algab igas uuringus kõrge tasemega kolesterooli mõõtmisest. Vajadusel leitakse mõõtmine interpoleerimisel aastaste vahemikega, peale esimest mõõtmist.
Sündmused	Uuritakse kolme võimalikku sündmust eraldi: 1. Aeg esimese esinemiseni sündmustest: <ul style="list-style-type: none"> • diagnoos I20-I25, I63 (v.a I63.6); • surm südame-veresoonkonna haigusesse. 2. Aeg esimese esinemiseni sündmustest: <ul style="list-style-type: none"> • diagnoos I21, I22, I63 (v.a I63.6); • surm südame-veresoonkonna haigusesse. 3. Aeg surmani südame-veresoonkonna haigusesse.	Samad, mis sihtkatses.
Huvipakkuvad põhjuslikud seosed	<i>intention-to-treat</i> mõju.	Sama, mis sihtkatses.

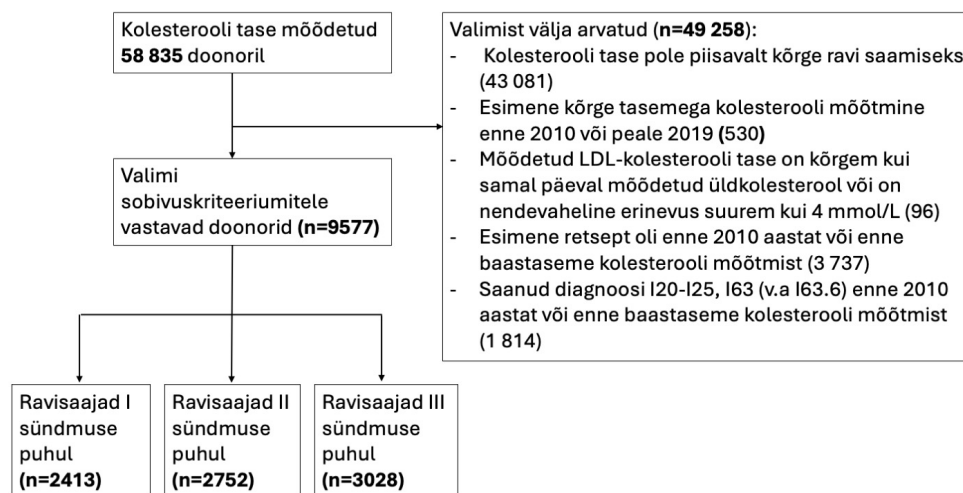
Ravistrateegiate puhul on määratud 1-aastane armuaeg, et kaasata uuringusse kõik isikud, kes on jälgimisaaja jooksul saanud ravi. Kui kavandaksime katse, kus strateegia nõuaks ravi kohest alustamist mõõtmise hetkel, hõlmaks katse strateegiaid, mida kliinilises praktikas harvemini esineb. Seega loetakse indiviid ravisaja katsetes, kui indiviidile on kuni ühe aasta jooksul peale jälgimisaaja algust kirjutatud välja kolesterooli alandav ravi.

Ajas muutuvate baastaseme tunnuste, nagu kehamassiindeks ja suitsetamine, väärtuseks võetakse igas katses jälgimisaaja algusele kõige lähemal olev väärtus. LDL-kolesterooli väärtused saadi lineaarse interpoleerimise teel, kus peale viimast mõõtmist jäid väärtused püsima viimase mõõtmise tasemele, nagu ka valimisse kaasamise kriteeriumites. Diagnoos peab olema registreeritud enne jälgimisaaja algust, et seda saaks arvestada segajana. Jälgimisaeg lõppeb tsenseerimisega, kui viimase andmete uuendamise hetkeks pole toimunud huvipakkuvat sündmust või surma, või kui jälgimisperiood on kestnud üle kümne aasta.

2.2.2 Valimi koostamine ja kirjeldav statistika

Geenivaramu doonoritest koostati valim vastavalt protokollis määratud sobivuskriteeriumitele. Joonisel 5 on kujutatud täpsemad protseduurid doonorite leidmiseks, kes vastavad vähemalt ühe uuringu valimisse sobivuse kriteeriumitele.

Vähemalt ühel aastal panustab katsesse 9 577 doonorit, kellest 34% on mehed. Jooniselt 5 on näha ka ravisajate arvud nende doonorite seas. Enne esimest sündmust või tsenseerimist on retsepti saanute osakaal valimis 25,2%, teise sündmuse puhul 28,7% ning kolmanda sündmuse puhul 31,6%. Mediaan jälgimisaeg nende doonorite seas on 6,6 aastat ja keskmine vanus jälgimisaaja alguses 50 eluaastat. Mediaanaeg esimesest kõrge tasemega kolesterooli mõõtmisest kuni kolesterooli alandava ravi retseptini on 24 kuud, kuid mediaanaeg esimese kardiovaskulaarsündmuse esinemiseni on vähem kui 20 kuud. Esimest huvipakkuvat kardiovaskulaarsündmust esineb nende doonorite seas 1258 isikul, teist sündmust 469 isikul ja kolmandat

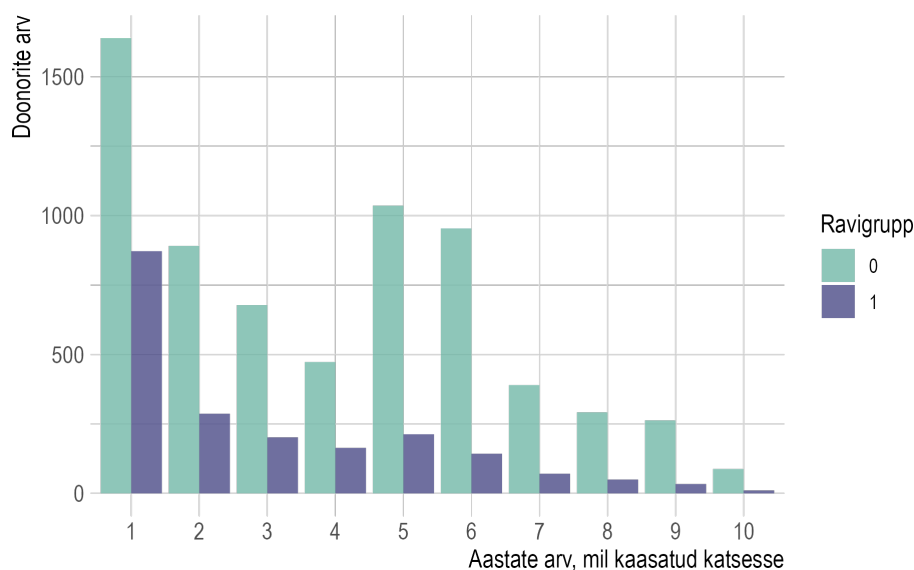


Joonis 5: Valimi moodustamise skeem koos vastavate valimimahtudega. Ravisaajad esimese, teise või kolmanda sündmuse puhul on vastavalt nende doonorite arv, kes enne esimese, teise või kolmanda sündmuse esinemist või tsenseerimist said kolesterooli alandava ravi retsepti.

sündmust 142 isikul.

Eelnevalt leitud doonoritele – kes osalevad katses vähemalt ühel aastal – määratakse, mitmes uuringus nad kokku osalevad, järgides valimisse sobivuse kriteeriume iga uuringu alguses. Kokku matkitakse 10 järjestikust sihtkatset vahemikus 2010-2019. Joonisel 6 on näidatud, mitmel aastal iga doonor uuringusse panustas esimese huvipakkuva sündmuse puhul.

Vastavalt valimisse sobivuse kriteeriumitele doonoreid uuringusse kaasates saame esimese sündmuse puhul kokku 31 550 isiku-uuringut, mis on valimisse kuuluvate doonorite arv üle koondatud uuringute (iga doonor panustab igasse kaasatud katsesse ühe isiku-uuringuga). Teise ja kolmanda huvipakkuva sündmuse puhul on isiku-uuringute arvud vastavalt 32 265 ja 32 374. Kõige enam doonoreid panustas ühte katsesse. Suurim ravisaajate arv on viimases katses algusega 2019 (341 ravisaajat ja 4909 mitte-saajat esimese huvipakkuva sündmuse puhul). Keskmiselt on



Joonis 6: Valimisse sobivuse kriteeriumitele vastavate doonorite jaotus aastate arvu järgi, mil nad katses osalesid, ravigrupi kaupa. Vaadatud on esimest huvipakkuvat sündmust. Joonisel 0 on ravi mitte-saajad ja 1 on ravisaajad.

esimese huvipakkuva sündmuse puhul iga aasta katsesse kaasatud 2982 doonorit ja keskmine ravisaajate arv igas katses on 173.

Tabelis 2 on välja toodud keskmised baastasemel mõõdetud kolesteroolitasemed valimis ja nende muutused peale kolesterooli alandava ravi alustamist.

Tabel 2: Kolesteroolitasemete keskmised muutused valimis peale ravi alustamist.

	Baastase	Std	Muutus	95% usaldusintervall
Üldkolesterool	6,81	0,72	-1,52	(-1,60; -1,43)
LDL-kolesterool	4,54	0,59	-1,34	(-1,43; -1,25)
HDL-kolesterool	1,54	0,46	0,019	(-0,05; 0,04)

Näeme ootuspärased langusi üld- ja LDL-kolesterooli keskmistes tasemetes peale ravi määramist. HDL-kolesterooli keskmine tase ei ole mõjutatud ravist, mis on samuti ootuspärane, kuna vaadeldavate kolesterooli alandavate ravimite mõju on suunatud spetsiifiliselt LDL-kolesteroolile.

2.3 Andmeanalüüs

Teostame analüüsi järjestikuste sihtkatsete matkmise meetodiga, nagu on selgitatud peatükis 1.3.3. Esimeses uuringus vaatame isikuid, kes vastavalt defineeritud ravistrateegiatele said ravi ja neid, kes kuuluvad alternatiivse ravistrateegia kohorti katses algusega 2010. Täpsemalt liigitatakse valimisse kuuluvad doonorid ravisaaajateks ja mittesaaajateks igal järjestikusel aastal vahemikus 2010-2019. Seliselt saame leida polügeense riskiskoori ja kolesterooli alandava ravi põhjusliku mõju südame-veresoonkonna haiguste riskile. Riskisuhete hinnangute leidmiseks koondatakse andmed üle kõikide uuringute ja sobitatakse andmetele Coxi mudel, mida kohandatakse baastaseme tunnustele. Usaldusintervallide leidmiseks kasutatakse 200 kordusega *bootstrap*-meetodit. Suuremahulised arvutused viidi läbi Tartu Ülikooli kõrgjõudlusega arvutusklastris.

Tabelis 3 on esitatud ravi kavatsuse mõju huvipakkuvate sündmuste riskile, mida on hinnatud Coxi mudeliga. Baastase on mitte-ravisaajad. Mudel on kohandatud tunnustele sugu, haridus, vanus, aasta, suitsetamine, LDL-kolesterool, diabeet, KMI, kõrgvererõhktõbi, lipoproteiiniainevahetuse häire, krooniline neerupuudulikkus ja kas inimene kuulub esimesse või teise EGV kohorti. Viimasele tunnusele on kohandatud, kuna nendes kahes kohordis on märkimisväärselt erinev südame-veresoonkonnahaiguste risk ja üldsuresus. Andmetes olevatest isiku-uuringutest, mis on üle kõigi katsete koondatud doonorite arv valimis, jääb esimese sündmuse puhul mudelist välja 13 459 isiku-uuringut mõne või mitme eelneva tunnuse puudumise tõttu. Teise ja kolmanda sündmuse puhul veel umbes 300 isiku-uuringut lisaks sellele.

Ravi kavatsuse mõju kardiovaskulaarsündmuste riskile ilmneb vaid teise huvipakkuva sündmuse juures. Esimese huvipakkuva sündmuse alla kuulub ka näiteks rinnaangiin (järsult tekkiv rinnatagune valu), mis oma olemuselt ei pruugi olla tõsine. Samuti hõlmavad need diagnoosid tüsistusi, kroonilist südame-isheemiatõve ja muid südame-isheemiatõve vorme. Need on spetsiifilised diagnoosid, mille pu-

Tabel 3: Ravi kavatsuse mõju huvipakkuvate sündmuste riskile.

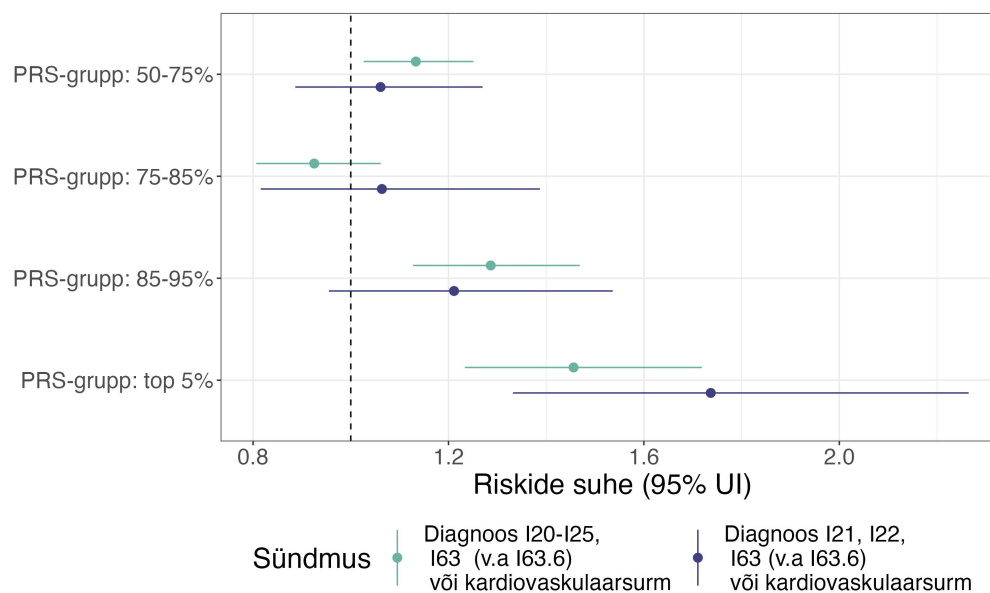
Sündmus	$\exp(\hat{\beta})$	95%- usaldusintervall	p -väärtus
Diagnoos I20-I25, I63 (v.a I63.6) või surm südame-veresoonkonna haigusesse	0,93	(0,77; 1,13)	0,4692
Diagnoos I21, I22, I63 (v.a I63.6) või surm südame-veresoonkonna haigusesse	0,56	(0,41; 0,78)	0,0006
Surm südame-veresoonkonna haigusesse	1,15	(0,81; 1,65)	0,4258

hul saab meditsiiniliselt põhjendada, miks kolesterooli alandav ravi nende riski ei vähenda.

Sündmusena ainult surma südame-veresoonkonna haigustesse vaadates ei tule samuti välja riski alandav mõju. Paneme tähele, et märkasime juba valimi koostamise käigus, et selle sündmuse esinemiste arv on väike. Seega on võimalik, et antud uuringus ei suudetud tuvastada ravi kavatsuse mõju kardiovaskulaarsurma riskile ebapiisava võimsuse tõttu.

Teise huvipakkuva sündmuse riski vähendab ravi kavatsus 0,56 korda (95% UI 0,41 kuni 0,78). See tähendab, et ravi määramisega on seotud 44% väiksem risk müokardi- ja peaaajuinfarktide ning kardiovaskulaarsurma esinemiseks võrreldes kontrollrühmaga. Äge müokardiinfarkt, mida see sündmus sisaldab, on südameisheemiatõve raskeim vorm.

Järgnevalt uurime südame-veresoonkonnahaiguste polügeense riskiskoori mõju huvipakkuvate sündmuste riskile. Hinnangud on saadud Coxi mudelist, mida on kohandatud tunnustele sugu, vanus, ravistrateegia, aasta ning indikaator doonori kuulumise kohta esimesse või teise EGV kohorti. Kuna PRS sõltub ainult isiku geneetilise profiilist ja geenid määratakse juhuvalikuga juba meioosi käigus, ei kaasne PRS-i mõju hindamisega samu baastaseme segajaid, nagu raviefekti hindamisel.



Joonis 7: Südame-veresoonkonnahaiguste riskiskoori mõju kahele esimesele huvipakkuvale sündmusele.

Andmetes olevatest isiku-uuringutest jääb mudelitest välja 167 isiku-uuringut puuduvate väärtuste tõttu. Joonisel 7 on toodud polügeense riskiskoori mõju kahe esimese huvipakkuva sündmuse riskile. Baastase on alla mediaani riskiskoori kategooria.

Polügeenne riskiskoor on jagatud viide kategooriasse vastavalt kvartiilidele. Kaks alumist kvartiili on kombineeritud ja ülemine kvartiil on jagatud kaheks väiksemaks protsentiiliks. PRS-i mõju kardiovaskulaarsuremuse riskile ei õnnestunud tuvastada. Jooniselt 7 näeme, et südame-veresoonkonnahaiguste riskiskooril on mõju kahe esimese huvipakkuva kardiovaskulaarsündmuse esinemisele. Isikutel riskiskoori kõrgeimas kategoorias on vastavalt 1,46 (95% UI 1,23 kuni 1,72) ja 1,74 (95% UI 1,34 kuni 2,25) korda suurem risk esimese ja teise huvipakkuva sündmuse esinemiseks kui riskiskooriga alla mediaani olevatel isikutel. Esimese huvipakkuva sündmuse puhul on alla mediaani jäävatest isikutest kõrgem risk veel riskiskoori kategooriates 50-75% ja 85-95%. Nende kategooriatega kaasneb vastavalt 13,3% (95% UI

1,8-26,1%) ja 28,6% (95% UI 14,1-44,9%) suurem risk esimese huvipakkuva sündmuse esinemiseks. Sellega oleme valideerinud südame-veresoonkonnahaiguste riskiskoori mõju kardiovaskulaarsündmuste esinemise riskile. Ühegi sündmuse puhul ei õnnestunud tuvastada ravi ja riskiskoori vahelist koosmõju. Siiski tasub uurida, kas koosmõju puudub ka PRS-i ja ravi järgimise vahel. Saanud sisukaid tulemusi ravi määramise analüüsist on järgmine samm teostada analüüs ravi järgimise mõju kohta.

Kokkuvõte

Sihtkatse matkimine nõuab hoolikat kaalutlemist, et vältida võimalikke nihkeid ja tuvastada põhjuslikke seoseid huvipakkuvate tunnuste vahel. Vaatlusandmete puhul on oluline mõelda, kuidas matkida võimalikult täpselt randomiseeritud katset, nagu see oleks teostatud kõikide ressursside olemasolul. Selleks on vaja panna paika sihtkatse põhilised komponendid ja kirjeldada, kuidas neid vaatlusandmete peal rakendada.

Töös on välja töötatud sihtkatse protokoll, mille abil koostatakse sobiv valim ja andmestik, et hinnata polügeense riskiskoori ja kolesterooli alandava ravi mõju südame-veresoonkonna haiguste riskile, kasutades Tartu Ülikooli Eesti geenivaramu andmeid. Sihtkatse protokollid viiakse vastavusse vaatlusandmetel järjestikuste sihtkatsete matkimise analoogiaga. Saadud andmetel hinnatakse kolesterooli alandava ravi määramise ja polügeense riskiskoori mõju huvipakkuvate kardiovaskulaarsündmuste riskile. Leitud hinnangud põhinevad Coxi mudelil saadud riskisuhetel ja nende robustsetel dispersiooni hinnangutel.

Kolesterooli alandava ravi määramise ja polügeense riskiskoori mõju surma riskile kardiovaskulaarhaigustesse ei õnnestunud kindlaks teha, mis võib olla tingitud andmete piiratud võimsusest. Esimese huvipakkuva sündmuse riskile ei leitud samuti kolesterooli alandava ravi mõju. Sündmus oli defineeritud nii, et hõlmas ka kergemaid vereringeelundite haigusi ja seetõttu polnud ka kliiniliselt ootamatu, et ravi ei avaldanud statistiliselt olulist mõju. Ravi määramisega oli seotud aga 44% väiksem risk müokardi- ja peaaajuinfarktide ning kardiovaskulaarsurma esinemiseks ehk huvipakkuvatest sündmustest kõige raskemaid südame-isheemiatõve haiguste vorme sisaldava sündmuse esinemiseks. Polügeense riskiskoori mõju südame veresoonkonna haiguste riskile avaldus kahe esimese huvipakkuva sündmuse puhul.

Töö edasiarendusena võiks järjestikuseid sihtkatseid defineerida lühemate ajavahe-
mike kaupa, näiteks igakuiselt. Kuna sellise uuringu teostamist piirab arvutusjõudlus, võib kaaluda kõigi ravisaajate kaasamist uuringusse ja juhuvalimi võtmist ravi

mitte-saajatest. See lähenemine võimaldaks defineerida lühema armuaja järjestikuste katsete teostamisel ja siiski kaasata teavet kõigi ravisaajate kohta. Samuti on sihtkatsete matkimisel antud teema jaoks veel mitmeid rakendatavaid meetodeid, mida selles töös ei käsitletud. Järjestikuste sihtkatsete matkimine on üks neist. Loomulik edasiarendus oleks uurida veel ravimi järgimise mõju südame-veresoonkonna haiguste riskile ja selle võimalikku koosmõju polügeense riskiskooriga, mis ei ilmenud ravi määramise analüüsis.

Kasutatud allikad

- Cardiovascular diseases*. World Health Organisation (2021). URL: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)) (vaadatud 29.12.2023).
- Collett, David (2023). *Modelling survival data in medical research*. Chapman ja Hall/CRC.
- Eesti geenivaramu*. Tartu Ülikool (2021). URL: <https://genomics.ut.ee/et/sisu/eesti-geenivaramu-0> (vaadatud 01.05.2024).
- Efron, Bradley ja Robert Tibshirani (1985). “The bootstrap method for assessing statistical accuracy”. *Behaviormetrika* 12.17, lk. 1–35.
- Hernán, Miguel A ja James M Robins (2010). *Causal inference*.
- (2016). “Using big data to emulate a target trial when a randomized trial is not available”. *American Journal of Epidemiology* 183.8, lk. 758–764.
- Hernán, Miguel A, Brian C Sauer, Sonia Hernández-Díaz, Robert Platt ja Ian Shrier (2016). “Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses”. *Journal of Clinical Epidemiology* 79, lk. 70–75.
- Keogh, Ruth H, Jon Michael Gran, Shaun R Seaman, Gwyneth Davies ja Stijn Vansteelandt (2023). “Causal inference in survival analysis using longitudinal observational data: Sequential trials and marginal structural models”. *Statistics in Medicine* 42.13, lk. 2191–2225.
- Kutcher, Stephen A, James M Brophy, Hailey R Banack, Jay S Kaufman ja Michelle Samuel (2021). “Emulating a randomised controlled trial with observational data: an introduction to the target trial framework”. *Canadian Journal of Cardiology* 37.9, lk. 1365–1377.
- Läll, Kristi (2019). “Risk scores and their predictive ability for common complex diseases”. Doktoritöö. Tartu Ülikool.

- Lodi, Sara, Andrew Phillips, Jens Lundgren, Roger Logan, Shweta Sharma, Stephen R Cole, Abdel Babiker, Matthew Law, Haitao Chu, Dana Byrne *et al.* (2019). “Effect estimates in randomized trials and observational studies: comparing apples with apples”. *American Journal of Epidemiology* 188.8, lk. 1569–1577.
- Patel, Aniruddh P, Minxian Wang, Yunfeng Ruan, Satoshi Koyama, Shoa L Clarke, Xiong Yang, Catherine Tcheandjieu, Saaket Agrawal, Akl C Fahed, Patrick T Ellinor *et al.* (2023). “A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease”. *Nature Medicine* 29.7, lk. 1793–1803.
- Sotsiaalministeerium* (2023). URL: <https://rhk.sm.ee> (vaadatud 13.01.2024).
- Surmad. Statistikaamet* (2024). URL: <https://www.stat.ee/et/avastatistikat/valdkonnad/rahvastik/surmad> (vaadatud 17.05.2024).
- Üldinfo. Tartu Ülikooli Eesti geenivaramu* (2023). URL: <https://geenidonor.ee/geenivaramu> (vaadatud 01.05.2024).

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Saskia Kuusk,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Sihtkatse matkimine: polügeense riskiskoori ja kolesterooli alandava ravi mõju südame-veresoonkonna haiguste riskile,“ mille juhendajad on Krista Fischer ja Lili Milani, 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.

Saskia Kuusk

22.05.2024