TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Kirke Valt

Andmepõhiste ravitrajektooride konstrueerimise meetodite süstemaatiline analüüs

Bakalaureusetöö (9 EAP)

Juhendaja:
Markus Haug, MSc

Tartu 2025

Andmepõhiste ravitrajektooride konstrueerimise meetodite süstemaatiline analüüs

Lühikokkuvõte:

Tervishoid on viimastel aastatel tegemas digipööret, kus järjest suuremat rolli mängib terviseandmete süsteemne kogumine, töötlemine ja analüüs. Käesolevas bakalaureusetöös analüüsiti andmepõhiste ravitrajektooride konstrueerimise meetodeid. Uurimistöö eesmärk oli kaardistada olemasolevad lahendused, hinnata nende reprodutseeritavust ning neid võimalusel implementeerida HPC-SAPU serveris RITA-MAITT. Läbi vaadati 622 uurimust, millest kirjandusanalüüsis leiti 15 tööd, reprodutseeritavuseanalüüsi läbisid kolm ning reaalselt implementeeriti nendest kaks. Töö tulemused näitavad, et kuigi ravitrajektooride modelleerimiseks on välja pakutud mitmeid innovaatilisi lahendusi, piirab nende ülekantavust ebapiisav dokumentatsioon ja pakettide ning puudulik uuringu koodide ühilduvus uute tarkvara versioonidega.

Võtmesõnad: terviseinformaatika, ravitrajektoorid

CERCS: B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

A Systematic Review of Methods for Mapping Health Trajectories

Abstract:

In recent years health care has undergone a digital revolution where an increasingly central role is played by the systematic collection and analysis of health data. This Bachelor's thesis investigates methods for constructing data-driven health trajectories. The objective was to map existing solutions, evaluate their reproducibility and implement them in the RITA-MAITT project on the HPC-SAPU server. A total of 622 articles were examined, from which 15 passed the literary analysis. Reproducibility analysis resulted in three remaining articles, out of which two were implemented. The findings highlight that even though many innovative solutions have been suggested for health trajectory mapping, their transferability is constrained by lack of documentation and compatibility issues with new versions of analytical software.

Keywords: health informatics, health trajectories

CERCS: B110 Bioinformatics, medicine informatics, biomathematics, biometrics

Sisukord

Sissejuhatus

Tervishoid on viimastel aastatel läbi tegemas ulatuslikku digipööret, kus järjest suuremat rolli mängivad terviseandmete süsteemne kogumine, töötlemine ja analüüs. Elektroonilised haiguslood, kantavad seadmed ning terviseinfosüsteemid salvestavad patsiendi kohta ajas järjestatud sündmusi luues seeläbi unikaalse ravikäigu ehk ravitrajektoori. Ravitrajektooride modelleerimine ja analüüsimine on muutumas järjest olulisemaks suunaks nii andmeteaduses kui meditsiinis, kuna see võimaldab tuvastada mustreid, hinnata sekkumiste mõju ja toetada personaalmeditsiini arengut (nt Siggaard et al., 2020). Eestis, kus terviseandmed on laialdaselt digitaliseeritud, on kvaliteetsetel ravitrajektooride kaardistamise mudelitel rohkelt potentsiaali parandada raviteekondade mõistmist, tervishoiu tõhusust ja personaalmeditsiini kvaliteeti.

Kuigi päriselu andmetel (ingl *real-world data*) on lahendatud mitmeid meditsiinilisi probleeme, on siiani probleemiks lahenduste lokaalsus ja erinevad andmeformaadid (Kent et al., 2021; Cave et al., 2019). Sõltuvus *ad-hoc* lahendustest nii programmeerimisel kui uuringute konfigureerimisel vähendab lahenduste efektiivsust ja skaleeritavust, mis omakorda raskendab uuringu või meetodi reprodutseerimist kolmandate osapoolte poolt (Wang et al., 2022). Sellest tulenevalt on tähtis, et erinevate publitseeritud mudelite valideerimiseks jagataks ka töövooga seotud koodi ja piisavalt detailset seletust, kuidas uuringut reprodutseerida. Sellised kitsaskohad tõstavad vajadust ühtsete andmeformaatide ja metoodikate järele, mis võimaldaksid tõhusamat koostööd eri asutuste ja riikide vahel.

Töö eesmärk on uurida ja dokumenteerida kirjanduses levinud ravitrajektooride koostamiseks kasutatud töövooge ning analüüsida nende kasutatavust Eesti tervishoiuandmetel neid võimalusel reprodutseerides. Töö tulemuseks on süstemaatiline kirjanduse ülevaade teaduslikes artiklites avaldatud ravitrajektooride koostamise meetoditest, nende reprodutseeritavuse hindamine ning praktiline rakendamine Eesti terviseandmetel.

Uurimus koosneb kolmest põhiosast. Esmalt kirjeldatakse metoodikat, selgitades töö käigus kasutatud analüüsi- ja hindamisetappe. Teises peatükis viiakse läbi põhjalik kirjandusanalüüs, mille eesmärk on süstemaatiliselt kirjeldada ja võrrelda teadusartiklite lähenemisviise ning hinnata nende metoodilist mitmekesisust ja praktilist rakendatavust. Kolmandas peatükis keskendutakse sobivate töövoogude reprodutseerimisele Eesti terviseandmetel, et hinnata nende praktilist teostatavust ja kohandatavust kohaliku andmekeskkonnaga.

1. Analüüsi metoodika

Selles peatükis kirjeldatakse töös kasutatud meetodeid ning andmete käsitlust. Käesolev töö põhineb süstemaatilise ülevaate meetodil, mis on teadustööde analüüsimiseks ja sünteesimiseks struktureeritud ja põhjalik lähenemine. Süstemaatiline ülevaade hõlmab eelnevalt kindlaks määratud protokolli järgimist, et tagada töö kirjutamisel maksimaalne objektiivsus. See protsess koosneb mitmest etapist: uurimisküsimuse täpne määratlemine, asjakohaste allikate otsimine ja filtreerimine, valitud allikate analüüs ning tulemuste süntees (Kohli 2020).

Töö käigus rakendatakse materjalide kogumiseks *snowballing*-tehnikat, mis on laialt levinud meetod süstemaatiliste ülevaadete kirjutamiseks (Mourão et al., 2020). Meetodi abil täiustatakse algselt tuvastatud allikate hulka, uurides viidatud või viitavaid allikaid. See lähenemine võimaldab laiendada ja süvendada olemasoleva materjali katvust, aidates leida olulisi ja varem märkamata jäänud allikaid. Kombineerides süstemaatilise ülevaate ranguse ja *snowballing*-tehnika paindlikkuse, tagatakse analüüsi terviklikkus ning selleks vajalike andmete mitmekülgne käsitlus.

1.1 *Snowballing*-meetod

1.1.1 Määratlus

Snowballing on üks tõhusamaid akadeemilise kirjanduse otsimise meetodeid (Wohlin et al., 2022). Tehnika eesmärk on suurendada käsitlevate allikate hulka, tuginedes algselt tuvastatud allikate viidetele. See meetod on tõestanud oma efektiivsust süstemaatilistes analüüsides, kuna algsete otsingustrateegiate kaudu leitavad artiklid ei pruugi alati hõlmata kogu teemaga seotud relevantset kirjandust. Meetod võimaldab uurimust edukalt laiendada ning tagada, et olulised vaatenurgad ei jääks töö käigus puudutamata. Tehnika suurimaks eeliseks on selle paindlikkus ja põhjalikkus ning peamiseks miinuseks on ajakulukus, kuna suure hulga teadustööde ning -artiklite läbitöötamine ja hindamine nõuab süvenemist, arusaamist ja aega.

Teema „ravitrajektooride konstrueerimise meetodid“ hõlmab suurt hulka erinevaid lähenemisviise ja uuringuid, mis võivad omada relevantsust mitmes teadusvaldkonnas. Oma olemuselt ei soosi meetod ainult leida vajalike allikaid vaid soodustab ka läbitöötleja arusaama teemavaldkonnast tervikuna. Antud tehnika kasutamine aitab põhjalikumalt kaardistada kõiki asjakohaseid meetodeid ja lähenemisviise, leides selliseid uuringuid, mis ei pruugi tavapärastes andmebaasides silma paista (Wohlin et al., 2014).

1.1.2 Protsess

Siinse uurimistöö metoodiline lähenemine tugineb süstemaatilisele ja iteratiivsele protsessile, mis kujutab endast asjakohaste teadustööde kasutatud kirjanduse läbitöötamist, otsides nende seast aktuaalseid töid ning seda tööülesannete käiku pidevalt jätkates, kuni läbitöödeldud materjal on tarvilik ja piisav käesoleva teadustöö kirjutamiseks.

Protsessi käigus kasutatakse kombineeritult nii tagasi- kui ka edasivaatavat *snowballing*-meetodit. Tagasivaatava meetodi rakendamine tähendab, et vaadatakse üle leitud artikli viidete loend, et tuvastada varasemalt kirjutatud käsitletava teemaga seotud uuringuid. Meetod omab erilist rolli, kui soovitakse leida uuringuid, mis on andnud aluse uuritavale teemale. Tagasivaatav *snowballing* aitab tõhusalt leida varasemaid olulisi viidatud allikaid, mida pelgalt andmebaasiotsinguga ei pruugi leida. Edasivaatava meetodi rakendamisel kasutati platvorme nagu Google Scholar, et leida teadustöid, kus on vastavat artiklit refereeritud. See meetod on eriti kasulik, kui uuritakse valdkondi, kus teadus areneb kiiresti ja on oluline kajastada hilisemaid ning kaasaegsemaid uuringuid. Edasivaatav meetod täiendab tagasivaatavat meetodit, võimaldades leida värskemaid uurimusi, millele on käesolev artikkel sisendit andnud. Reeglina käivad need meetodid käsikäes ning kasutatakse praktikas sageli koos, et tagada ülevaate terviklikkus.

Materjalide kaasamise või välistamise kriteeriumid seatakse tulenevalt töö eesmärkidest. Kirjanduse lisamisel kasutatakse samu kriteeriume nagu mugavusvalimisse tööde kaasamisel (Mellor, 2022). Iga potentsiaalse materjali puhul kaalutakse selle panust uurimuse terviklikkuse ja kvaliteedi tõstmisese seisukohast. Hindamisprotsess hõlmab mitmeid olulisi aspekte. Analüüsitakse materjali asjakohasust, kuivõrd käsitletav uuring haakub käesoleva tööga, usaldusväärsust, metoodilist kvaliteeti ning ajalist relevantsust. Samuti viiakse loetud materjalide seas läbi kriitiline analüüs, mille eesmärk on tõsta esile teadustöö uuringu fookusega kõige tihedamalt seotud teadmised ja luua alus järgmisteks uurimisetappideks. Välistamiskriteeriumiteks seatakse samuti selged põhimõtted. Kõrvale jäetakse allikad, mille metoodika, tulemus või eetika on küsitavad, mis kordavad olemasolevaid andmeid sisuliste panusteta ning mis on üldistatud või arvamuspõhised. Kõrvale jäetakse ka tööd, mis olid avaldatud enne 2010. aastat, kuna nende sisu ei ole enam ajakohane ega kajasta praeguseid teadusuuringute arengusuundi.

Snowballing-meetodi rakendamiseks vajaliku mugavusvalimi koostamisel võeti arvesse erinevaid ravitrajektooride tüüpe (Joonis 2). See tähendab, et iga trajektoori tüübi jaoks valiti

üks esinduslik teaduslik artikkel. Selle tulemusena osutusid valituks töö juhendajale teadaolevad teadusartiklid Reps jt (Reps et al., 2018), Künnapuu jt (Künnapuu et al., 2022) ning Haug jt (Haug et al., 2024).

1.2 Ravitrajektooride käsitlemise meetodid

Käesolevas töös keskendutakse konkreetselt ravi, mitte haigustrajektooride käsitlemisele. Peamine erinevus on, et ravitrajektoor koosneb kõikidest meditsiinilistest sekkumistest nagu ravimine või sümptomite kontrolli alla saamine. Seevastu haigustrajektoor, mis on osa ravitrajektoorist, kirjeldab peamiselt vaid tüsistusi ja nende vahelisi üleminekuid..

Ravitrajektooride käsitlemiseks vajalikud algandmed tulenevad enamasti patsiendi terviseandmetest, sealhulgas haiguslugudest, meditsiiniliste testide tulemustest, arstivisiitide- ning hospitaliseerimisandmetest (Barker, Scherer 2019). Lisaks mängivad kohati ka rolli kvalitatiivsed andmed patsientide kogemusest ning ravi mõjust nende igapäevaelule.

Trajektooride konstrueerimiseks eraldatakse algandmed vastavalt aja- ning teemakohasele fookusele. Lisaks eristatakse ka andmeid, mis kajastavad otseselt ravi efektiivsust, kõrvalmõjusid ning patsiendi subjektiivset rahulolu. Ravitrajektoorid esitatakse andmetena mitmel erineval viisil sõltuvalt uurimiseesmärgist. Enim kasutatud vormid trajektooride esitamiseks on näiteks aja- ja sündmuspõhised graafikud, statistilised mudelid (Abbasi 2024). Aja- ja sündmuspõhised graafikud näitavad ravi algust, olulisi sekkumisi ja tulemusi konkreetsel ajateljel ehk oma olemuselt on visualiseeritud ravi algus- ja lõpp-punktid koos patsiendi ravi kulgemisega. Statistilisi mudeleid kasutatakse suuremahuliste terviseandmete analüüsimiseks ja prognoosimustrite tuvastamiseks, et leida korrelatsiooni ravi kestuse, tulemuse ja vajalike meditsiiniliste testide vahel (Jensen et al., 2014; Beck et al., 2016).

Süstemaatilisse ülevaatesse kaasati artikleid, mis käsitlesid mudeleid, mis põhinesid läbipaistvatel (ingl *white-box*) mudelitel nagu Markovi mudelid, otsustuspuud ja reeglipõhised lähenemised. Töösse ei kaasatud mitte tõlgendatavaid (ingl *black-box*) meetodeid käsitlevaid artikleid nagu sügavad närvivõrgud. See kitsendus on seatud töö juhendaja poolt.

Käesolevas teadustöös keskendutakse konkreetselt vaatlusandmetel põhinevatele ravitrajektooridele. See on ajendatud eesmärgist analüüsida ja reprodutseerida valitud meetodeid Tartu Ülikooli terviseinformaatika töögrupi hoiustatud andmetel. Töö praktilises osas reprodutseeritakse leitud ravitrajektooride konstrueerimise töövooge RITA-MAITT projekti andmetel, mis koosneb 10% eestlaste terviseandmetel (Oja et al., 2023). Kogu andmete

töötlemine toimub käesoleva töö raames *High-Performance Computing Centre Sensitive data Analysis Platform* (edaspidi HPC SAPU) serveris *Observational Medical Outcomes Partnership Common Data Model* (edaspidi OMOP CDM) andmetel, et tagada patsiendiandmete turvalisus ja konfidentsiaalsus. Töö andmetega viidi läbi vastavalt TÜ eetikakomitee ja Eesti bioeetika ja inimuuringute nõukogu lubadele (load nr 300/T-23 ja 1.1-12/3088) ning projektide TEM-TA72 ja PRG1844 raames. Projekt TEM-TA72 on rahastatud Euroopa Liidu ja kaasrahastatud Haridus- ja Teadusministeeriumi poolt. Projekt PRG1844 on rahastatud Eesti Teadusagentuuri poolt.

1.3 Artiklite reprodutseeritavuse hindamine

Selle peatüki aluseks on võetud Tatman jt artikkel (Tatman et al., 2018), kus on välja toodud etapid masinõppe uurimuste reprodutseeritavuse hindamiseks, ning neid on kohandatud vastavalt töö spetsiifikale ja eesmärkidele. Varasemalt koostatud etapiviisilist lähenemist on kasutatud juhisena töö ülesehitamisel, kuid etappide sisu on vajadusel muudetud, täiendatud või täpsustatud, et need vastaksid paremini uuritava teema eripäradele ja konkreetsele kontekstile.

Uurimistööde reprodutseeritavuse hindamiseks kasutatakse viit etappi, mis on kohandatud lähtuvalt ravitrajektoore kaardistavate tööde spetsiifikast. Iga etapi puhul määratakse artiklile väärtus, täieliku olemasolu korral linnuke, puudumise korral rist ning kui kriteeriumid on täidetud puudujääkidega märgitakse väärtuseks “osaliselt” (Joonis 3).

Esmalt hinnatakse, kas uurimuse autorid on avaldanud kasutatud koodi ning kas see on kättesaadav usaldusväärsel ja avalikul platvormil (nt GitHub). Täiendavalt vaadeldakse, kas on loodud spetsiaalne tarkvarapakett, mis võimaldab metoodikat rakendada ka väljaspool algse uurimuse konteksti. Teisena pööratakse tähelepanu sellele, kas töös kasutatud andmestik on avalikult kättesaadav, standardiseeritud formaadis (nt OMOP CDM) või vähemalt piisavalt hästi dokumenteeritud. Hinnatakse, kas autorid on kirjeldanud andmete päritolu, struktuuri, puhastamise ja eeltöötluse protsessi ning andmestikus sisalduvate kliiniliste mõistete määratlemist või kas on esitatud näidislõik andmetest. Töö väärtuseks hinnati “osaliselt”, kui esines andmestruktuuri kirjeldusi, kuid mitte piisaval määral, et neid täpselt replikeerida. Kolmandana analüüsitakse, kas töö tulemusi on võimalik reaalsete sammudena korrata. Oluline on, kas on kaasas täpsed juhised analüüsi läbiviimiseks ning kas koodifailid on struktureeritud viisil, mis võimaldab reprodutseerida tulemused minimaalsete muudatustega. Kriteerium märgiti osaliselt täidetuks, kui mõned etapid olid hästi dokumenteeritud, kuid teised (nt

andmetöötlus või visualiseerimine) olid kas puudu, ebaselged või põhjendamata. Neljandaks kriteeriumiks hinnatakse, kuivõrd detailselt on kirjeldatud kasutatud meetodid ja modelleerimisloogika. Tähelepanu pööratakse sellele, kas ravitrajektooride defineerimine, ajavahemike valik, algoritmide tööpõhimõtted ning kasutatud parameetrid on selgelt esitatud ja põhjendatud. Metoodiline läbipaistvus märgiti osaliselt täidetuks juhul, kui töö andis üldise ülevaate, kuid ei kirjeldanud piisavalt detailselt ravitrajektooride loomise loogikat või ei põhjendanud kasutatud parameetreid. Viiendaks kriteeriumiks on reprodutseerimise kogemus ehk töö varasem taasesitamine mõnes teises andmestikus või kontekstis. Kuigi see ei ole reprodutseeritavuse eeltingimus, viitab selline korduv kasutamine töö praktilisele rakendatavusele ja usaldusväärsusele. Ideaalis peaks kvaliteetne teadustöö vastama kõigile eelnevatele kriteeriumidele ning olema hõlpsasti reprodutseeritav ka erinevates kontekstides.

2. Ülevaade

Käesolevas peatükis antakse põhjalik ülevaade kasutatud metoodikast ning esitatakse tulemused, mis on saadud metoodika rakendamisel. Lisaks analüüsitakse teadusartikleid, mille uurimisküsimused ja meetodid sarnanevad käesoleva töö omadega.

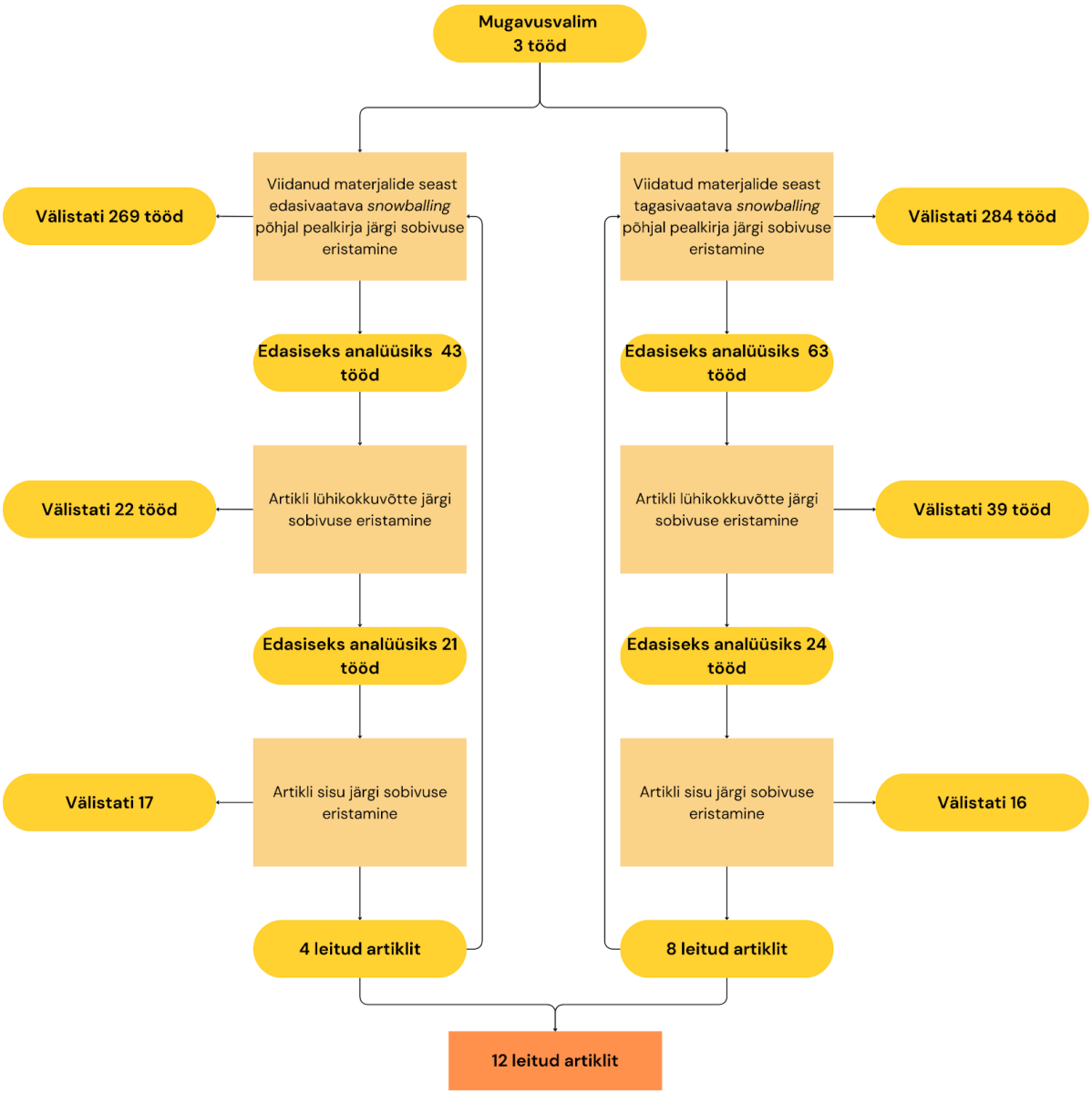
2.1 Aktuaalsete artiklite analüüs

Ravitrajektooride loomine ja analüüsimine on valdkond, mis on oma olulisuse poolest kogumas aina enam kõlapinda nii andmeteaduses kui ka meditsiinis (Kerexta-Sarriegi et al., 2024). Selle peatüki eesmärk on esitada põhjalik ülevaade teadusartiklitest ja uuringutest, mis käsitlevad ravitrajektooride loomist ja modelleerimist. Aktuaalsete artiklite leidmiseks ja süstemaatiliseks analüüsiks kasutati eelmainitud *snowballing*-meetodit.

Tööde filtreerimiseks on kasutatud *forward-* ja *backward-snowballing* meetodit ehk võeti ette vastavalt kas mugavusvalimi tööde viidatud kirjanduse loetelu või mugavusvalimi töödele viidanud kirjanduse loetelu ning hakati välja filtreerima käesolevasse töösse sobivaid materjale. Mugavusvalim valiti teadusuuringute seast, mis olid eelkõige seotud Tartu Ülikoolis tehtud uurimistöödega, samas tõdeti, et valimi sobivust hinnati ka selle järgi, et valitud teadusartiklid käsitleksid ka erinevaid trajektooride modelleerimise lähenemisviise. See tagas mitmekesise ülevaate valdkonna uurimistöödest, võimaldades võrrelda erinevaid metoodilisi suundi ja nende rakendusi praktikas.

Süstemaatilise kirjanduse ülevaate koostamise töö käik on konkreetselt esitatud joonisel 1. Analüüsis vaadati läbi koos mugavusvalimiga 662 tööd. Esmalt tehti valik pealkirja põhjal. Kõige lihtsam oli välja jätta artiklid, mis olid süstemaatilised analüüsid või rangelt uurimuslikud teadustööd. Selles etapis välistati 85,5% kõikidest töödest ning edasiseks analüüsiks jäi 109 artiklit. Järgnevalt uuriti täpsemalt tööde lühikokkuvõtteid, millest jäeti edasi uurimiseks alles 45 artiklit, mis käsitlesid ravitrajektoore. Viimase etapi käigus selgitati välja töö sobivus süvitsi sisu analüüsides. Selles etapis jäeti analüüsist välja 33 tööd, mis käsitlesid ravimustrite analüüsi, kuid tuginesid peamiselt eelnevalt avaldatud kirjandusele või agregeeritud andmetele. Kuna nende eesmärgiks ei olnud patsienditasemel terviseandmete põhjal individuaalsete raviprotsesside modelleerimine, ei vastanud need käesoleval töö fookusele. Viimases etapis jäeti ka välja tööd, mis olid avaldatud enne 2010. aastat, kuna nende sisu ei olnud enam ajakohane ega kajastanud praeguseid teadusuuringute arengusuundi. Analüüsist leiti lisaks mugavusvalimile 12 artiklit, mis vastasid loodud kriteeriumitele,

moodustades kõikidest analüüsitud töödest kõigest 1,8% ning kõikidest vähemalt pealkirja põhjal relevantseks hinnatud töödest ligikaudu 11%.



Joonis 1. Teadustööde analüüs *snowballing*-meetodi abil.

Skeemil on välja toodud kõik erinevad leitud artiklid, st joonisel 1 kujutatud numbrid esindavad unikaalseid artikleid.

Leitud artiklid esitati kokkuvõtvalt tabelina (Tabel 1). Tabeli lihtsustamiseks määrati ravitrajektooride käsitlustüüpidele tähised A, B või C. Ravitrajektoori tüübid on visualiseeritud joonisel 2, kus S tähistab sündmust ehk igasugust meditsiinilist sekkumist. Trajektoori tüüp A tähistab trajektoore, mis põhinevad sündmuspõhisel lähenemisel (ingl *Binary Event Trajectories*). Selline lähenemine tähendab, et analüüsis käsitletakse mitmeid erinevaid

meditsiinilisi sekkumisi, kus iga patsiendi kohta registreeritakse binaarne väärtus – kas vastav sündmus esines (1) või ei esinenud (0). Tähtis on märkida, et antud tüüpi ravitrajektoorid ei arvesta ajalise järjestuse ega haiguse progresseerumisega, vaid keskenduvad üksnes sündmuse esinemise faktile. See metoodika on kasulik teatud tüüpi mustrianalüüsides, kuid on üsna piiratud võimekusega haiguskulu dünaamika või ravi mõju hindamisel. Seda meetodit kasutatakse peamiselt masinõppe ja andmepõhise meditsiiniuuringute kontekstis, kus aeg ei ole esmane prioriteet (Beaney et al., 2024). Meetodi eeliseks on lihtsus ja efektiivsus suurte andmekogumite töötlemisel, kuid see võib kaotada olulist teavet, mis on seotud haiguste ajatelje või raviprotsessi muutustega (Piao et al., 2024).

A – Sündmuspõhine ravitrajektoor

	S1	S2	S3	S4	S5	S6	...
#1 Patsient	0	1	1	0	0	0	
#2 Patsient	1	0	1	0	1	1	

B – Esmaesinemisel põhinev ravitrajektoor

	S1	S2	S3	S3	...
#1 Patsient	NA	21.06.2024	NA	20.04.2024	
#2 Patsient	13.02.2022	4.06.2024	NA	10.06.2024	

C – Rekurrentne ravitrajektoor

	S1	S2	S3	...
#1 Patsient	List(2020, 2005, 2005)	NA	List(2006)	
#2 Patsient	NA	List(2023, 2022, 2019)	List(2022, 2021)	

Joonis 2. Ravitrajektooride tüübid.

Trajektoori tüüp B tähistab esmaesinemisel põhinevat ravitrajektoori (ingl *First-Occurrence Temporal Trajectory*). Erinevalt A-kategooria ravitrajektooridest mängib selles käsitlusviisis aeg keskset rolli. Analüüs keskendub meditsiiniliste sekkumiste ajateljele, registreerides üksnes esimese esinemise iga diagnoosi puhul. See tähendab, et kui patsient saab hilisemas elus uuesti juba varem esinenud diagnoosi, siis käesoleva metoodika raames seda ei arvestata.

Esmakordselt esinenud diagnoosidest moodustatakse paarid, mille vahelisi seoseid hinnatakse nende esinemissageduse ja treeningandmetes esineva relevantsuse asemel, et tuvastada potentsiaalsed järelduvad seosed meditsiiniliste sekkumiste vahel. See meetod võimaldab analüüsida haigusprotsesside kronoloogiat ja tuvastada tüüpilisi järjestusmustreid erinevate patoloogiate esinemisel.

Trajektoori tüüp C tähistab oma olemuselt kõige keerukamaid rekurrentseid ravitrajektoore (ingl *Occurrence Temporal*). Erinevalt B-kategooria ravitrajektooridest, mis on oma olemuselt lineaarsed ja keskenduvad üksnes esmastele esinemistele, võimaldavad C-tüüpi trajektoorid arvesse võtta ka korduvaid diagnoose, mistõttu võivad need sisaldada tsüklilisi mustreid. See tähendab, et patsiendi haiguslugu ei ole pelgalt järjestikune sündmuste jada, vaid võib endast kujutada keerukamat dünaamilist protsessi, kus sama diagnoos võib aja jooksul korduda. Üks levinumaid C-tüüpi trajektooride modelleerimise meetodeid on Markovi mudel. Oma informatiivsuse ja struktuurilise keerukuse tõttu on C-tüüpi trajektoorid analüütilises ja kliinilises kontekstis kõige olulisemad ning potentsiaalselt kõige väärtuslikumad haiguste prognoosimisel ja ravistrateegiate kujundamisel.

Tabel 1. Analüüsiks valituks osutunud tööd

Autorid	**Aasta**	**Andmed**	**Valim**	**Meetod(id)**	**Trajektoori eesmärk**	**Tüüp**
Reps et al.	2018	Diagnoosid	~4 mln	OMOP CDM-põhine andmeanalüüs, masinõppepõhised ennustusmudelid, mudeli valideerimine	Standardiseeritud ennustus-raamistiku väljatöötamine ja rakendamine	A
Künnapuu et al.	2022	Diagnoosid ravimid,	~10 mln	OMOP CDM- põhine andmeanalüüs, Ajaliste kliiniliste sündmuste järjestuste tuvastamine	Terviseandmete analüüsi kirjeldamine ja visualiseerimine	B
Haug et al.	2024	Diagnoosid, ravimid, protseduurid	47 163	Markovi modelleerimine, tervisemajanduse analüüs	Kulutõhususe hindamine	C
Weng et al.	2017	Diagnoosid, protseduurid, ravimid	378 256	Masinõppe algoritmid (random forest, logistic regression, gradient boosting, neural networks)	Kardiovaskulaarse riski ennustamine	A
Siggaard et al.	2020	Diagnoosid	7,2 mln	Disease trajectory browser andmemahtude analüüsimiseks,	Haiguste progresseerumine erinevates	B

				statistilised analüüsid, masinõpe	inimrühmades ja ajas, tuvastada riskigrupid ja haiguse kulgu mõjutavad tegurid	
Jensen et al.	2014	Diagnoosid	6,2 mln	Masinõpe, statistilised mudelid	Haiguste arengu mõistmine ja ennustamine	B
Beck et al.	2016	Diagnoosid	6,6 mln	Andmeanalüüs, masinõpe, statistilised mudelid	Sepsisega patsientide suremuse ennustamine	B
Hu et al.	2019	Diagnoosid	6,9 mln	Masinõpe, statistiline analüüs	Määrata vähieelsete seisundite arenguteed ja seost vähi tekkega	B
Romeu	2020	Diagnoosid, ravimid, protseduurid	-	Markovi mudel	Covid-19 patsientide ellujäämise tõenäosuste hindamine ja haiguse kulgemine	C
Soper et al.	2020	Diagnoosid, protseduurid	1,7 mln	Peidetud Markovi mudel	Emakakaelavähi sõeluuringu tõhususe parandamine, ülediagnoosimise vähendamine	C
Uhry et al.	2010	Haigusstaadiumid, staadiumite üleminekud	-	Mitme-staadiumilised Markovi mudelid	Sõeluuringu mõju hindamine erinevates staadiumites, et optimeerida sõeluuringu strateegiaid	C
Yang et al.	2022	Diagnoosid, protseduurid, ravimid	-	Masinõpe, statistline modelleerimine, välise valideerimise analüüs	Riskipatsientide varajane tuvastamine ja ravi edendamine	A
Lin et al.	2022	Diagnoosid, ravimid, protseduurid, tüsistused	-	Masinõpe, statistiline modelleerimine	Operatsioonijärgsete tüsistuste ennetamine, riskipatsientide ennustamine	A
Hetland et al.	2024	Diagnoosid, ravimid	-	Masinõpe	Tuvastada infektsioonide risk ja ennustada neid	A
Bräuner et al.	2024	Patsiendi näitajad,	57 521	Masinõpe, statistiline modelleerimine	Ennustada operatsioonijärgset suremust, tuvastada riskitegurid	A

Läbiva analüüsi põhjal võib trajektooripõhised lähenemised jaotada kolme tüüpi vastavalt nende eesmärgile: A-tüüp keskendub eelkõige prognoosimisele, B-tüüp kirjeldamisele ja visualiseerimisele ning C-tüüp valdavalt tervisemajanduslike ja otsustusmudelite rakendamisele. Iga tüübi puhul rakendati erinevaid meetodikomplekte sõltuvalt eesmärgist ja andmete laadist.

A-tüüpi ravitrajektooride uuringutes oli domineerivaks lähenemiseks masinõpe – kasutati mitmeid algoritme nagu *random forest*, regressioonanalüüsid, *gradient boosting* ja närvivõrgud. Lisaks masinõppele kaasati sageli ka statistiline modelleerimine ning välise valideerimise meetodid, et tugevdada mudelite üldistatavust ja usaldusväärsust. Selliseid meetodeid kasutati peamiselt riskipatsientide tuvastamiseks ning erinevate ennustuste tegemiseks.

B-tüüpi uuringud keskendusid peamiselt haigusmustrite ja arengutrajektooride tuvastamisele ning visualiseerimisele. Nende puhul kasutati graafipõhiseid lähenemisi, kus graafi tippude ja servade leidmiseks rakendati statistilisi teste. Näiteks Künnapuu et al. artiklis kasutati OMOP CDM-põhist andmeanalüüsi koos ajalisel järjestusel põhinevate meetoditega sündmuste ühendamiseks. Oluline on märkida, et mitmed hilisemad B-tüübi tööd, nagu Siggaard et al. (2020), on üles ehitatud varasemate uuringute metoodikale, näiteks tugineb nende kasutatud *Disease Trajectory Browser* tööriist Jensen et al. (2014) poolt loodud metoodikale haiguste ajaliseks järjestamiseks ja trajektooride modelleerimiseks. Kuigi ka siin kasutati masinõpet, ei olnud see keskne tööriist, vaid pigem toetas mustrite leidmist.

C-tüüpi ravitrajektooride uuringutes rakendati ennekõike Markovi ahelaid. Need võimaldasid hinnata näiteks sõeluuringute efektiivsust, haiguse kulgu ajas või sekkumiste kulutõhusust. Terviseökonoomika mudelid olid selles tüübis domineerivad, viidates vajadusele simuleerida erinevaid stsenaariume ja hinnata nende mõju süsteemi tasandil. Ravitrajektooride koostamiseks kasutati nii spetsiifilise uuringu andmeid (Soper et al., 2020; Uhry et al., 2010) kui reeglipõhist lähenemist (Haug et al., 2024).

Joonistub selgelt välja, et trajektooritüüp määrab suurel määral kasutatavate meetodite valiku – prognoosimisel domineerib masinõpe, haigusmustrite uurimisel statistilised ja graafilised lähenemised ning otsustusanalüüsis Markovi modelleerimine. Samuti ilmneb, et mitmed uuringud põhinevad eelnevatel töödel ja olemasolevatel raamistikkudel, mis aitab kaasa meetodite standardiseerimisele ja tulemuste reprodutseeritavusele teaduskogukonnas.

Järgnevalt antakse ülevaade mõistetest nagu OMOP CDM ja Markovi ahelad, millele viidati korduvalt läbitöödeldud artiklites.

2.1.1 OMOP CDM

Observational Medical Outcomes Partnership (OMOP) *Common Data Model* (CDM) on standardiseeritud andmemudel, mida kasutatakse tervishoiu vaatlusuuringutes, et ühtlustada ja struktureerida erinevatest allikatest pärit terviseandmeid. See on siiani üks kõige terviklikumaid püüdlusi luua universaalne terviseandmete käsitlemise ja kajastamise süsteem. OMOP CDM-i eesmärk on võimaldada andmete ülekantavust, reprodutseeritavust ja võrreldavust eri tervishoiusüsteemide ja uurimisrühmade vahel. Mudeli eeliseks on selle lai rakendusala. See toetab mitmesuguseid terviseandmete tüüpe, sealhulgas diagnoose, protseduure, ravimeid ja laborianalüüse. Ühtlustatud andmestruktuur võimaldab erinevate andmeallikate võrdlemist ja loob aluse standardiseeritud analüüsitööriistade väljatöötamiseks. (*Understanding OMOP Basics*, 2024)

2.1.2 Markovi ahela mudelid

Markovi ahela mudelid pakuvad informatiivset ja kvaliteetset raamistikku ravitrajektooride modelleerimiseks, võimaldades kirjeldada patsiendi seisundi muutumist ajas tõenäosuslike üleminekute kaudu. Enamjaolt jagunevad mudelid diskreetse aja Markovi mudeliks, pideva aja Markovi mudeliks ja peidetud Markovi mudeliks. Diskreetse aja Markovi mudelites toimub üleminek ühest seisundist teise kindlaksmääratud ajaühikute kaupa, näiteks päevade nädalate või kuude lõikes. See määratakse vastavalt uuringu iseloomule, ning seda tüüpi Markovi mudel on sobilik eelkõige kliinilisteks uuringuteks ja ravistrateegiate hindamiseks (Gautam, 2009). Pideva aja Markovi mudelid seevastu võimaldavad seisundimuutusi igal ajahetkel, modelleerides üleminekuid eksponentsiaalse jaotusega, mis on kasulik haigusprotsesside dünaamika detailsemaks analüüsiks (Whitt, 2014). Peidetud Markovi mudelid laiendavad klassikalist Markovi raamistikku, käsitledes juhtumeid, kus patsiendi tegelik seisund ei ole otseselt vaadeldav, vaid seda tuleb järeldada kaudsetest vaatlustest (Rabiner, Juang, 1986).

2.2 Sarnaste artiklite analüüs

Sarnasel teemal on ka varasemalt süstemaatilisi kirjanduse ülevaateid koostatud. Alljärgnevas peatükis antakse ülevaade kahest tööst, mis sarnanevad oma olemuselt käesolevale tööle, ning selgitatakse käesoleva ning käsitletava artikli erinevusi.

2.2.1 Haigustrajektooride süstemaatiline analüüs

Autorite Jørgensen, Haue, Placido, Hjaltelin ja Brunak artiklis "*Disease Trajectories from Healthcare Data: Methodologies, Key Results, and Future Perspectives*" (Jørgensen et al., 2024) keskendutakse haiguste trajektooride kaardistamisele tervishoiuandmete põhjal, kasutades erinevaid analüüsimeetodeid. Uuritakse haiguste progresseerumise mustreid, riskitegurite mõju ja interventsioonide tulemuslikkust. Kasutatakse meetodeid nagu statistilised masinõppemeetodid, võrgustikuanalüüs ning heterogeensete andmete integreerimise strateegiad. Lisaks kasutatakse võrgustikuanalüüsi haiguste vaheliste suhete mõistmiseks ning selleks, et ühendada geneetilisi, sotsiaalmajanduslikke ja kliinilisi andmeid.

Artikli eesmärgiks on pakkuda süsteemset ülevaadet meetoditest, mis võimaldavad analüüsida ja modelleerida haiguste kulgemist, et toetada personaliseeritud ravi ja tervishoiusüsteemide optimaalsemat toimimist. Artikli motivatsioon tuleneb vajadusest mõista haiguste arengut ja soovi pakkuda prognoosimudelite kaudu efektiivsemaid ravivõimalusi.

Käesolev töö erineb antud artiklist teemakäsitluse poolest. Uuritav artikkel keskendub haigustrajektooridele ehk haiguste progresseerumisele ja nende vahelistele mustritele. Käesolev töö paneb rõhku ravitrajektooridele , mille eesmärgiks on täiendavalt raviotsuste, protsesside mustrite ja tulemuste kirjeldamine. Viimane lähenemine seab esile patsiendikesksed aspektid. Haigustrajektoorid seevastu on osa ravitrajektooridest ning seetõttu esitlevad ka oluliselt kitsamat pilti. Uuritavas artiklis rõhutatakse reprodutseeritavuse seisukohalt olulisi probleeme, mis tulenevad tervishoiuandmete konfidentsiaalsusest ja piiratud juurdepääsust analüüsimeetoditele. Ravitrajektooride uurimisel lisandub täiendav keerukus, mis tuleneb erinevate osapoolte ja ravietappide mitmekesisusest ning patsiendikesksusest.

2.2.2 Kulutõhususe analüüs Markovi mudelitega

Teine sarnase ülesehituse ja fookusega töö on "*Simple but not simpler: a systematic review of Markov models for economic evaluation of cervical cancer screening*" (Viscondi et al., 2018), kus autorid analüüsivad süsteemselt Markovi mudeleid, mida on kasutatud emakakaelavähi sõeluuringute kulutõhususe hindamiseks. Analüüsitakse 36 teadusartiklit, keskendudes eelkõige mudelite struktuurilisele ülesehitusele, sisendandmete allikatele ning mudelite valideerimise viisidele. Lisaks hindavad nad ka sarnaselt käesolevale tööle, kuivõrd läbipaistvad ning korduvkasutatavad on kirjeldatud mudelid ning kui hästi need vastavad nõuetele tervishoiupoliitika kontekstis.

Artikli eesmärk on hinnata, kuivõrd mudelid vastavad hea modelleerimistava põhimõtetele, keskendudes metoodilisele kvaliteedile ja praktilisele rakendatavusele tervishoiupoliitika kujundamisel. Markovi ahela mudel on siiani üks parimaid viise dünaamilise ravitrajektoori kaardistamiseks.

Käesolev töö lähtub sarnasest loogikast, pakkudes süsteemset ülevaadet, kuid peamises fookuses on ravitrajektoorid üldiselt. Kasutatakse sarnaseid meetodeid kirjanduse uurimisel ning andmete esitamisel. Mõlemas töös on keskne roll tervishoiuga seotud modelleerimismeetodite süsteemsel analüüsil. Uuritav töö on rangelt kvantitatiivne ning suunatud kulutõhususe hindamisele, samas käesolev töö keskendub kasutatud meetodite kirjeldamisele ja reprodutseerimise hindamisele.

3. Reprodutseeritavuse hindamine ja sobivate artiklite implementeerimine

Käesolevas peatükis analüüsitakse ning hinnatakse leitud artiklite reprodutseeritavust ning sobivad artiklid ka reprodutseeritakse.

3.1 Tööde reprodutseeritavus

Eelnevalt valiti analüüsitav aktuaalne kirjandus välja kasutades edasi- ja tagasivaatavat *snowballing*-meetodit. Käesolevas peatükis hinnatakse nende tööde reprodutseeritavust lähtudes metoodika peatükis kirjeldatud viiest etapist. Iga leitud töö puhul analüüsitakse eraldi, mil määral vastavad need esitatud etappidele. Analüüsi tulemused on kokku võetud joonisel 3. Hindamine keskendub sellele, kas ja kuidas on artiklites dokumenteeritud koodi kättesaadavus, andmestiku kirjeldus, kasutatud metoodika detailid ning juhised analüüsi läbiviimiseks, et hinnata tööde reprodutseeritavuse taset.

Analüüsitud artiklitel on enim puudujääke andmete kirjelduses. Kirjeldus on tihtipeale ebapiisav ning detailsus on jäänud pigem tahaplaanile. Kõikides töödes kasutatud andmed ei olnud avalikult kättesaadavad, mistõttu on negatiivse mõjuga reprodutseerimise seisukohalt, sest reprodutseerimisel on kriitilise tähtsusega andmete piisav kirjeldamine. Andmete kirjelduse piisavuse hindamiseks kasutati järgmisi kriteeriume: vaadeldi, kas töö sisaldas näidisandmestikku või kas andmeid kirjeldati nii detailselt, et üksnes tekstipõhise info põhjal oli võimalik andmestik töödelda mudelile sobivasse vormi. Sellisel juhul märgiti vastavasse lahtrisse linnuke. Kui aga andmete kirjelduses esines puudujääke või ebaselgust, mille tõttu ei olnud võimalik andmestikku täielikult ette valmistada, märgiti lahtrisse rist. Töödel, millel esineb pakett, on teoreetiliselt võimalik koodist vajalik andmete esitusviis välja selgitada, kuid see on ajakulukas, ebaproduktiivne ning ei pruugi lõplikult vastata koodi oodatud kriteeriumitele.

ARTIKLID	AVALDATUD KOOD	PIISAV ANDMETE KIRJELDUS	JUHISED	LÄBIPAISTEV METOODIKA	ON REPRODUTSEERITUD
Reps et al.	✔	✔	✔	✔	✔
Künnapuu et al.	✔	✔	✔	✔	✔
Haug et al.	✔	✔	✔	✔	✔
Weng et al.	✘	✔	osaliselt	✔	✘
Siggard et al.	✘	✔	osaliselt	✔	✘
Jensen et al.	✘	✔	osaliselt	✔	✘
Beck et al.	✘	osaliselt	osaliselt	✔	✘
Hu et al.	✘	osaliselt	osaliselt	✔	✘
Romeu	✘	✘	osaliselt	✔	✘
Soper et al.	✘	osaliselt	osaliselt	✔	✘
Uhry et al.	✘	✘	✘	osaliselt	✘
Lin et al.	✘	osaliselt	osaliselt	osaliselt	✘
Hetland et al.	✘	osaliselt	osaliselt	osaliselt	✘
Bräuner et al.	✘	osaliselt	osaliselt	osaliselt	✘

Joonis 3. Tööde reprodutseeritavuse hindamine.

Edukalt ning kvaliteetselt oli seevastu kajastatud töö metoodikad, enamasti oli detailselt kirjeldatud kasutatud meetodeid, analüüsietappe ning põhjendusi valitud lähenemisviiside kohta. See võimaldas hästi mõista töö ülesehitust ning hinnata selle teaduslikku rangust. Samas jäi liiga sageli mulje, et tekst on teadlikult või tahtmatult keeruliselt kirjeldatud – metoodika kirjeldustes kasutati komplitseeritud või liigselt tehnilist sõnastust, mis raskendas arusaamist. Soovituslikult võiksid autorid pöörata rohkem tähelepanu sellele, et metoodika kirjeldused ning üldse kirjutatud tekstid oleksid mitte ainult täpsed, vaid ka selged ja loogiliselt üles ehitatud, vältides liigseid tehnilisi detaile või keerukaid väljendeid seal, kus lihtsamast seletusest piisaks – eeldusel, et vajalik detailsus on esitatud koodina või lisamaterjalides (Gebru et al., 2021). Arusaadav ja loogiline tekst nii metoodika kui ka andmete kirjelduse kohta on

reprodutseeritavuse seisukohalt võtmetähtsusega, tagades üheti mõistetavuse ning usaldusväärse edasirakendamise.

Mõnes analüüsitud töös oli märgata, et viidates OMOP CDM-i kasutamisele jäädi andmete kirjelduse ja metoodika esitamisega pealiskaudseks. Kuigi OMOP-standard loob ühtse raamistiku ja toetab andmete struktureeritust, ei vabasta see autoreid kohustusest selgitada täpselt milliseid andmevälju, muutujaid ja valikuid konkreetses analüüsis kasutati. OMOP-i mainimine üksi ei taga, et lugeja või reprodutseerija saab andmestiku põhjal aru, kuidas see konkreetse uurimistöö kontekstis rakendus või milliseid eeldusi ja piiranguid autorid oma analüüsis arvestasid. Samas pakuvad just sellised standardiseeritud andmemudelid ja tööriistad olulist võimalust parandada uuringute läbipaistvust, reprodutseeritavust ja edasikantavust. OMOP-i ja teiste sarnaste platvormide kasvav kasutus, sealhulgas EHDENi ja DARWIN EU projektide kaudu Euroopas ning OHDSI globaalses koostöös, võib tulevikus oluliselt tõsta tervisandmete analüüsi kvaliteeti, eeldusel et neid rakendatakse teadlikult ja piisava detailsusega, kaasates ka majanduslike ja kulutõhususuuringute standardiseeritud käsitlusi (Haug et al., 2024).

Kuigi detailsed juhised analüüsi läbiviimiseks ei ole tingimata vajalikud reprodutseeritavuse tagamiseks juhul, kui kood ja andmestik on täielikult ning korrektselt dokumenteeritud, täidavad need siiski olulist toetavat funktsiooni. Konkreetsed ja struktureeritud juhised lihtsustavad oluliselt uurimuse taasesitamist, vähendades aja- ja ressursikulu, mida teised teadlased peavad töö mõistmiseks ja tulemuste kontrollimiseks tegema. Lisaks aitavad need vältida metodoloogilisi ebatäpsusi ning toetavad uurimistulemuste usaldusväärset valideerimist ja edasist rakendamist teadustöös.

Analüüsitud töödest oli reprodutseeritud vaid kolm tööd. Need olid ka ainsad tööd, mis vastasid ka kõikidele eelnevalt läbikäidud kriteeriumitele. Implementeerimiseks otseselt välistati tööd, millel ei olnud avaldatud rakendatud uuringu koodi.

3.2 Implementeeritavad tööd

Käesoleva peatüki eesmärgiks on hinnata, kuivõrd hästi on võimalik reprodutseerida teadusartikleid, mis vastasid eelnevalt määratletud valikukriteeriumitele. Reprodutseerimiseks valiti kolm artiklit, mis vastasid seatud tingimustele ning osutusid sobivaks ka tehnilise teostatavuse vaates. Reprodutseerimise läbiviimiseks kasutati projekti RITA-MAITT OMOP CDM formaadis andmeid HPC-SAPU serveris ning analüüsimiseks RStudio keskkonda.

Analüüsi läbiviimiseks kasutati tarkvara R versiooniga 4.4.0 ja RStudio versiooniga 2024.03.1+541.

3.2.1 Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data

Esimene artikkel, mis sobis eelmainitud kriteeriumitega arvestades, on "*Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data*" (Reps et al., 2018). Töö eesmärgiks on luua standardiseeritud ja avatud lähtekoodiga raamistik kliiniliste ennustusmudelite arendamiseks, jagamiseks ja reprodutseerimiseks OMOP CDM andmestiku põhjal.

Kuna uuritav töö tugines OHDSI loodud PLP (ingl *Patient-Level Prediction*) raamistikule ja OMOP CDM mudelile, siis esmapilgul tundus, et töö peaks olema hästi reprodutseeritav. Teoreetiliselt oleks reprodutseerimise protsess koosnema andmebaasi edukast ühendamisest, kohordi määramisest ning juba avaldatud koodi käivitamisest, kasutades PLP paketti.

Esmalt selgus, et kuigi artiklis oli viidatud PLP töövoogudele ja kohordiloogikale, ei olnud esmapilgul leitav konkreetselt artikli protsessi kajastav kood, mida reprodutseerida. Selle asemel oli olemas näidisprojekt *DepressionModels* (StudyProtocolSandbox, 2018), mis põhines vanemal PLP versioonil. Seal esitatud kood oli vormiliselt töökorras, kuid sõltus funktsioonidest, mis tänapäevases PLP versioonis enam ei eksisteeri. Kasutatud PLP versioon oli GitHubist kustutatud ning vajaolevad funktsioonid olid kas ümber nimetatud, eemaldatud või leidus nendele sarnanevaid, kuid mittesobivaid funktsioone.

Kuna PLP ja tema sõltuvused on pidevas arengus, on paljud kasutatud paketid vahepeal uuendatud ning mõni ka eemaldatud. Selle tulemusel ei olnud kood enam käivitatav. PLP sõltub mitmetest teistest OHDSI tööriistast ning nende kõigi õige versiooni rakendamine oli tehniliselt võimalik, kuid väga aeganõudev ja ebaproduktiivne protsess. Lisaks tekkis ka konflikt RStudio versiooniga, uuringus kasutatud PLP versioon nõudis sõltuvuspakette, mida uusim RStudio versioon ei toeta. Teoreetiline lahendus oleks olnud kogu keskkonna ajakohasele versioonile viimine, kuid see tundus ebapraktiline ning riskantne ülejäänud töö osas.

Kuigi artikkel oli teoreetiliselt reprodutseeritav, ei olnud selle tehniline pool enam ajakohane. Vaja oleks olnud spetsiifilist eraldatud keskkonda või vähemalt detailselt dokumenteeritud sõltuvuste nimekirja koos kasutatud R versiooniga. Joonistus konkreetselt välja, et

reprodutseeritavus ei tähenda ainult seda, et rakendatud kood ja pakett on kättesaadav, vaid ka seda, et seda on realistlik ja praktiline käivitada ilma liigse ajakuluta. Ideaalis võiks OHDSI-s või laiemalt teadusartiklite kõrval olla alati eraldi säilitatud konkreetne tööpakett ja keskkonna seadistusfail, mis võimaldaks töö kiiresti ja tõrgeteta taastoota.

3.2.2 Trajectories: a framework for detecting temporal clinical event sequences from health data standardized to the OMOP Common Data Model

Teiseks reprodutseerimiseks sobivaks osutus artikkel "*Trajectories: a framework for detecting temporal clinical event sequences from health data standardized to the OMOP Common Data Model*" (Künnapuu et al., 2022). Tegemist on metoodilise tööga, mis keskendub ajaliselt järjestatud kliiniliste sündmuste trajektooride tuvastamisel OMOP CDM andmetel.

Erinevalt esimese uuringu puhul tekkinud raskustest, osutus selle artikli reprodutseerimine oluliselt lihtsamaks. GitHubis oli esitatud Trajectories pakett (Trajectories, 2022), kus kättesaadavad failid olid hästi dokumenteeritud ja loogiliselt struktureeritud. Eriti kasulikuks osutus kaasasolev juhend (ingl *vignette*), mis kirjeldas täpselt, kuidas töövoog erinevates etappides kulgeb ning milliseid faile ja parameetreid saab kasutaja kohandada.

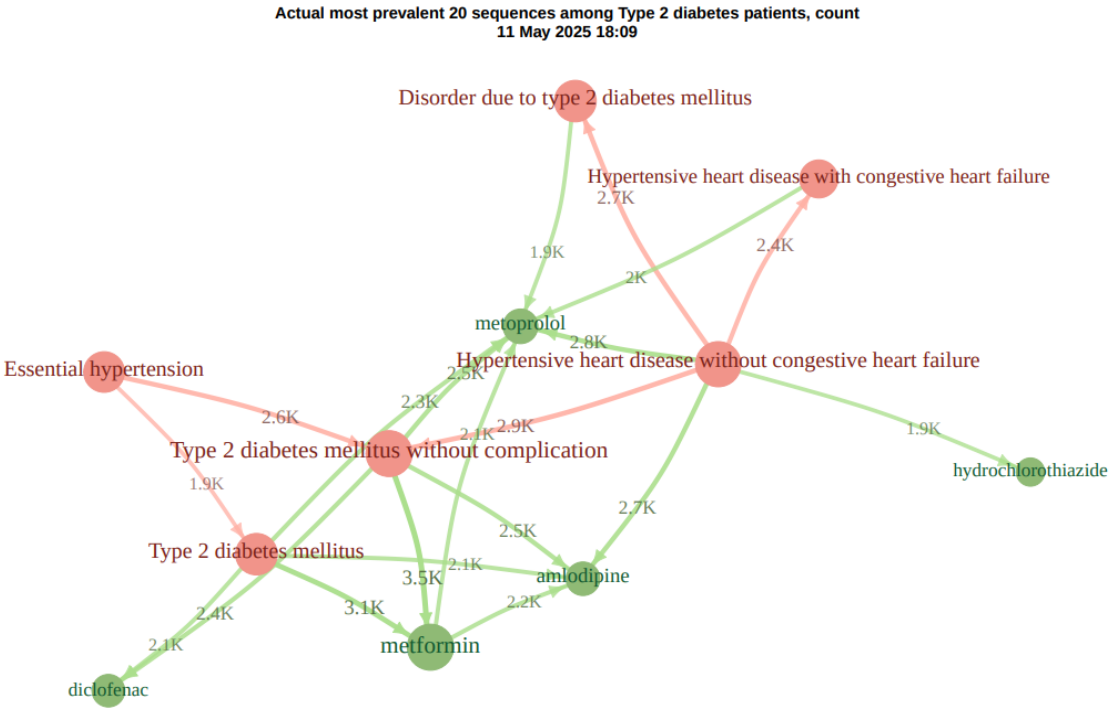
Reprodutseerimisprotsess jagunes kahte põhietappi:

Discovery mode – mille abil avastati sagedasemad sündmuste järgnevused (ingl *event pairs*) kindlaksmääratud kohortis.

Validation mode – kus avastatud mustreid valideeriti kas samas andmestikus (ingl *self-validation*) või alternatiivselt mõnes muus.

Trajectories pakett võimaldas neid samme selgelt eristada ning kontrollivalt käivitada. Avalikult olid olemas ka test-failid kohordimääratlustega, mis hõlbustas protsessi veelgi.

Reprodutseerimise käigus kasutati etteantud andmeid, mille alusel defineeriti 2. tüüpi diabeedi kohort. Kuna käesolevas töös oli kasutada ainult üks OMOP CDM andmestik, viidi valideerimine läbi sisemise jagamise teel (ingl *self-validation*), kus enne mustrite väljaselgitamist eraldati 25% andmetest valideerimiskogumiks. Kuigi selline lähenemine ei kinnita üldistatavust teistesse andmestikesse, võimaldab see hinnata mustri stabiilsust ja toimivust sama andmestiku sees. Analüüsis kasutati *discovery* ja *self-validation* funktsioone ning väljundina genereeriti sündmuste järgnevused csv-failina ning ka visuaalse trajektoorvõrguna, millel olid kujutatud nendest sagedasemad.



Joonis 4. 20 kõige sagedamini esineva sündmuse järgnevused.

Implementeerimisel tulemuseks saadud trajektoorvõrk (joonis 4) visualiseerib 20 kõige sagedasemat ajaliselt järjestatud sündmuste paari 2. tüüpi diabeedi patsientide seas. Graafilt on näha, kuidas seisundid (nt “Type 2 diabetes mellitus without complication”) viivad edasi spetsiifilistele ravimitele, nagu metformiin või amlodipiin. Sõlmede suurus peegeldab sündmuste esinemissagedust ja nooled indikeerivad ajas suunatud mustreid, kus kõik kuvatud seosed olid valideeritud andmestikusiseselt. Kahjuks on visuaalne trajektoorvõrk väljundina PDF-fail, seega ei saa seda interaktiivselt kohandada, et see loetavam oleks.

Käesolev töö on väga hästi reprodutseeritav, kui kasutajal on ligipääs OMOP CDM formaadis andmetele. Lisaks pakub pakett paindlikkust: kasutaja saab mugavalt kohandada ajavahemikku sündmuste vahel, täpsustada sündmuste tüüpe ning täpsustada trajektoorvõrgu koostamise reegleid. Artikli metoodika ei ole ainut teoreetiliselt avatud, vaid ka praktiliselt kergesti rakendatav ja kohandatav.

3.2.3 Markov modeling for cost-effectiveness using federated health data network

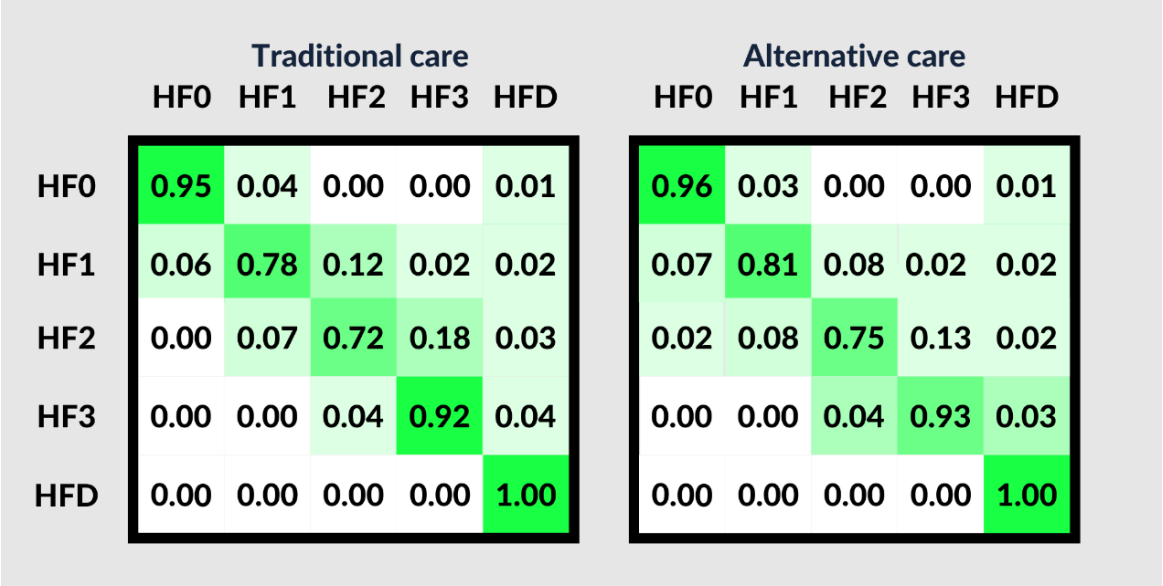
Kolmandaks ja viimaseks artikliks mida reprodutseeriti oli artikkel *"Markov modeling for cost-effectiveness using federated health data network"* (Haug et al. 2024). Artikkel keskendub südamepuudulikkusega (ingl *heart failure* edaspidi HF) patsientide kliiniliste ravitrajektooride ja ravikulude hindamiseks.

Artikli reprodutseerimine osutus väga hästi teostatavaks. Sarnaselt kõigile eelnevatele reprodutseeritud artiklitele, põhines ka see töö üles-ehitatud OMOP CDM andmetel. GitHubis esitatud töövoog oli detailselt kirjeldatud ning selgelt struktureeritud. Eriti väärtuslikuks osutus kaasasolev juhiste fail, kus oli täpselt defineeritud rakendatud paketid koos nende versioonidega, et vältida uuendatud pakettide mittesobivust uuringukoodiga.

Reprodutseerimise käigus tuli esmalt paigaldada kolm omavahel seotud paketti: *Cohort2Trajectory*, *HeartFailureTrajectoryCostStudy*, *TrajectoryMarkovAnalysis* (HealthInformaticsUT, 2023a, 2023b, 2023c), pidades silmas, et pakettide versioonid oleks õiged, ning täiendavad sõltuvused, mida paketid nõudsid. Pakettide ülesehitus oli moodulipõhine ja järgitav – igal osal oli oma funktsionaalne eesmärk, nt kohordiloome, trajektooride analüüs ja kulude modelleerimine. Oli ka spetsiifiliselt kommentaaridena välja toodud parameetrid, mida muuta või mida kindlalt muuta ei tohi.

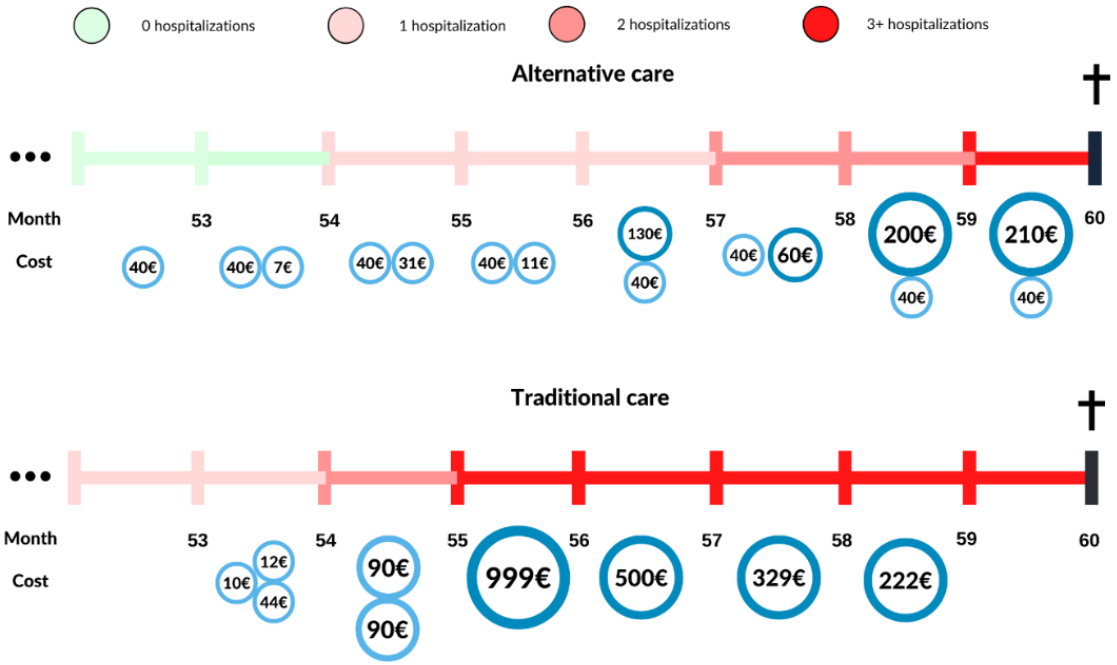
Pärast edukat paigaldust, andmebaasiga ühildamist ja integreeritud koodi jooksutamist genereeriti väljundfailid, mis visualiseerisid ravi tüüpide mõju patsientide kliinilistele trajektooridele ja kulumustritele.

Analüüs hõlmas 60 kuu pikkust ehk viieaastast vaatlusaega kuni patsiendi surmani. Patsiendid jaotati viite seisundisse (HF0-HF3 ja HFD) sõltuvalt hospitaliseerimise arvust vaatlusaja jooksul. Trajektoorid loodi kuu kaupa ning Markovi ahela üleminekutõenäosused arvutati suurima tõepära (ingl *maximum likelihood estimation*) meetodil. Kulud pärinesid OMOP CDM andmestiku *cost* tabelist.



Joonis 5. Markovi ahela üleminekumaatriksid HF patsientide seisundite vahel kahes ravigrupis.

Joonis 5 kuvab Markovi ahela üleminekumaatrikseid kahes ravigrupis: traditsiooniline ja alternatiivne ravi. Lahtrid näitavad patsiendi seisundite vahel liikumise tõenäosust. Seisund HF0 tähendab, et pole olnud ühtegi hospitalisatsiooni, HF1 ja HF2 tähistab vastavalt üks või kaks hospitalisatsiooni ning HF3 tähistab kolme või enam hospitalisatsiooni. Seisund HFD tähistab surma ehk on neelduv, mis tähendab, et sellest ei ole võimalik edasi liikuda. Joonisel 6 on visualiseeritud kahe patsiendirühma igakuiseid haiglakulusid. Joonis on illustratiivne ega põhine otseselt andmetel. Ringide suurus näitab kulutuste suurust ning värv illustreerib hospitaliseerimiste arvu. On selgelt näha, et traditsioonilise ravi puhul on kulud hüplikumad ja 55. kuul esineb väga suur kulu (999€), mis võib peegeldada näiteks korduvhospitaliseerimisi või intensiivravi, alternatiivravi korral on kulud madalamad ja stabiilsemad.



Joonis 6. Igakuised haiglakulud ja hospitalimiseerimiste arv traditsioonilise ja alternatiivravi korral.

Reprodutseerimine kulges üldjoontes sujuvalt, töövoog edenes ootuspäraselt ning dokumentatsiooni toel oli kõiki etappe võimalik täpselt korrata. Kindlasti kindlustas protsessi asjaolu, et tööd oli varasemalt implementeeritud samal andmebaasil, kuid teises serveris ning andmestiku versioonil. Käsitletud artikkel on eeskujulikult dokumenteeritud ja metoodiliselt läbimõeldud. Uuring on selgelt struktureeritud ning reprodutseerimine kinnitas, et pakutud metoodika on mitte ainult teaduslikult usaldusväärne, vaid ka praktiliselt rakendatav. Uuringu kasutajale, kellel eeldatavasti on ligipääs OMOP CDM andmetele, on kogu protsess tehtud arusaadavaks ja ühetimõistetavaks, mis näitab töö olulist väärtust edasistele uuringutele. Võimaliku kitsaskohana tooks välja nii Künnapuu jt kui ka Haug jt artikli puhul, et kuigi OMOP CDM pakub standardiseeritud raamistikku andmete esitamiseks ja analüüsimiseks, on selle rakendamine piiratud andmestike puhul, mis ei ole veel OMOP CDM vormingusse üle viidud.

Kokkuvõte

Käesolevas bakalaureusetöös viidi läbi süstemaatiline analüüs erinevate meetodite kohta, mida kasutatakse ravitrajektooride kaardistamiseks terviseandmete põhjal. Töö eesmärgiks oli kaardistada olemasolevad lahendused, hinnata nende reprodutseeritavust ning neid võimalusel implementeerida HPC-SAPU serveris RITA-MAITT projekti OMOP CDM andmestikul. Analüüsis kasutati *snowballing*-meetodit, mille abil valiti välja asjakohased teadusartiklid, milles kirjeldatakse erinevaid trajektooride konstrueerimise ja anlüüsimise metoodikaid.

Töö koosnes kolmest põhietapist. Esimeses kirjeldati detailselt rakendatud metoodikat ning seati analüüsile täpsed kriteeriumid. Selgitati *snowballing*-meetodi olemust ja efektiivsust ning rakendamisel hinnati materjali asjakohasust, usaldusväärsust, metoodilist kvaliteeti ning ajalist relevantsust. Kirjeldati detailselt ka ravitrajektooride käsitlemise meetodeid ning artiklite reprodutseeritavuse hindamisstrateegiat. Viimase puhul jälgiti viit seatud kriteeriumi. Hinnati, kas on avaldatud uuringu kood, kui jäljendatav on andmestik, kas ja kui selgelt on esitatud juhised reprodutseerimiseks, kui läbipaistev ja selge on kirjeldatud metoodika ning kas artiklit on juba varasemalt reprodutseeritud.

Teises etapis viidi läbi põhjalik kirjanduseanalüüs. Esiteks analüüsiti kirjandust rakendades *snowballing*-metoodikat võttes aluseks mugavusvalimi. Läbi vaadati 622 uurimust. Kriteeriumitele vastas lisaks mugavusvalimile 12 tööd, mis moodustas ligikaudu 11% relevantsest analüüsitud kirjandusest,. Leitud artiklid jagunesid kolme ravitrajektoori käsitlustüübi vahel: sündmuspõhine ravitrajektoor, esmaesinemisel põhinev ravitrajektoor ning rekurrentne ravitrajektoor. Leitud töödes esines enam esmaesinemisel põhinevate ravitrajektooride uuringuid. Lisaks kirjanduseanalüüsile avati ka kahe uurimistöö sisu, mis sarnanesid käesoleva tööga ning selgitati ja põhjendati käesoleva töö erinevust.

Kolmandas etapis hinnati kirjandusanalüüsist leitud artiklite reprodutseeritavust viiel seatud tingimustel ning sobivaid artikleid reprodutseeriti RITA-MAITT projekti OMOP CDM andmestikul. Peamised murekohad reprodutseeritavuse hindamise analüüsis olid andmete puudulik esitusviisis ja uuringukoodi puudumine. Loodud kriteeriumitele vastas kolm artiklit ning neist kaks olid praktiliselt ka edukalt reprodutseeritavad. Reprodutseerimisel olid peamisteks takistusteks aegunud pakettide versioonid.

Selleks, et publitseeritud töövood oleksid täielikult reprodutseeritavad, tuleb iga analüüsi samm esitada maksimaalse läbipaistvusega. Metoodika peab olema detailselt dokumenteeritud, sh

kasutatud algoritmid, tarkvaraversioonid ja kõik eel- ning järeltöötluse seadistused. Samuti on oluline lisada selge kirjeldus kasutatud andmetest kui need ei ole universaalses formaadis ning ka kättesaadav lähtekood koos täpse versiooni info ja sõltuvuste loeteluga.

Viited

Abbasi, K. (2024). How to model life's trajectory from major diagnosis to death. BMJ, 384, q568. https://www.bmj.com/content/384/bmj.q568 (5.12.2024).

Beaney, T., Jha, S., Alaa, A., Smith, A., Clarke, J., Woodcock, T., Majeed, A., Aylin, P., & Barahona, M. (2024). Comparing natural language processing representations of coded disease sequences for prediction in electronic health records. Journal of the American Medical Informatics Association, 31(7), 1451–1462. https://doi.org/10.1093/jamia/ocae091 (14.05.2025).

Beck, M. K., Jensen, A. B., Nielsen, A. B., Perner, A., Moseley, P. L., & Brunak, S. (2016). Diagnosis trajectories of prior multi-morbidity predict sepsis mortality. Scientific Reports, 6(1), 36624. https://doi.org/10.1038/srep36624 (17.02.2025).

Bräuner, K. B., Tsouchnika, A., Mashkoor, M., Williams, R., Rosen, A. W., Hartwig, M. F. S., Bulut, M., Dohrn, N., Rijnbeek, P., & Gögenur, I. (2024). Prediction of 30-day, 90-day, and 1-year mortality after colorectal cancer surgery using a data-driven approach. International Journal of Colorectal Disease, 39(1), 31. https://doi.org/10.1007/s00384-024-04607-w (31.03.2025).

Cave, A., Kurz, X., & Arlett, P. (2019). Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe. Clinical Pharmacology and Therapeutics, 106(1), 36–39. https://doi.org/10.1002/cpt.1426 (14.05.2025).

Gautam, N. (2011). Definition and Examples of DTMCs. J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, & J. C. Smith, Wiley Encyclopedia of Operations Research and Management Science (1. tr). Wiley. https://doi.org/10.1002/9780470400531.eorms0237 (31.03.2025).

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., III, H. D., & Crawford, K. (2021). Datasheets for Datasets (No. arXiv:1803.09010). arXiv. https://doi.org/10.48550/arXiv.1803.09010 (13.05.2025).

Haug, M., Oja, M., Pajusalu, M., Mooses, K., Reisberg, S., Vilo, J., Giménez, A. F., Falconer, T., Danilović, A., Maljkovic, F., Dawoud, D., & Kolde, R. (2024). Markov modeling for cost-effectiveness using federated health data network. Journal of the

American Medical Informatics Association: JAMIA, 31(5), 1093–1101. https://doi.org/10.1093/jamia/ocae044 (03.03.2025).

HealthInformaticsUT/Cohort2Trajectory. (2023a). GitHub. Retrieved May 15, 2025, from https://github.com/HealthInformaticsUT/Cohort2Trajectory/releases/tag/v1.1.3 (15.05.2025).

HealthInformaticsUT/HeartFailureTrajectoryCostStudy. (2023b). [R]. Health Informatics Lab, University of Tartu. https://github.com/HealthInformaticsUT/HeartFailureTrajectoryCostStudy (Original work published 2022) (15.05.2025).

HealthInformaticsUT/TrajectoryMarkovAnalysis. (2023c). GitHub. Retrieved May 15, 2025, from https://github.com/HealthInformaticsUT/TrajectoryMarkovAnalysis/releases/tag/v1.0.5 (15.05.2025).

Hetland, M. L., Strangfeld, A., Bonfanti, G., Soudis, D., Deuring, J. J., & Edwards, R. A. (2024). Machine learning prediction and explanatory models of serious infections in patients with rheumatoid arthritis treated with tofacitinib. Arthritis Research & Therapy, 26(1), 153. https://doi.org/10.1186/s13075-024-03376-9 (31.03.2025).

Hu, J. X., Helleberg, M., Jensen, A. B., Brunak, S., & Lundgren, J. (2019). A Large-Cohort, Longitudinal Study Determines Precancer Disease Routes across Different Cancer Types. Cancer Research, 79(4), 864–872. https://doi.org/10.1158/0008-5472.CAN-18-1677 (24.02.2025).

Jensen, A. B., Moseley, P. L., Oprea, T. I., Ellesøe, S. G., Eriksson, R., Schmock, H., Jensen, P. B., Jensen, L. J., & Brunak, S. (2014). Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nature Communications, 5(1), 4022. https://doi.org/10.1038/ncomms5022 (17.02.2025).

Jørgensen, I. F., Haue, A. D., Placido, D., Hjaltelin, J. X., & Brunak, S. (2024). Disease Trajectories from Healthcare Data: Methodologies, Key Results, and Future Perspectives. Annual Review of Biomedical Data Science (Kd 7, Number Volume 7, 2024, lk 251–

276). Annual Reviews. https://doi.org/10.1146/annurev-biodatasci-110123-041001 (3.02.2025).

Kent, S., Burn, E., Dawoud, D., Jonsson, P., Østby, J. T., Hughes, N., Rijnbeek, P., & Bouvy, J. C. (2021). Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment. PharmacoEconomics, 39(3), 275–285. https://doi.org/10.1007/s40273-020-00981-9 (14.05.2025).

Kerexeta-Sarriegi, J., García-Navarro, T., Rollan-Martinez-Herrera, M., Larburu, N., Espejo-Mambié, M. D., Beristain Iraola, A., & Graña, M. (2024). Analysing disease trajectories in a cohort of 71,849 Patients: A visual analytics and statistical approach. International Journal of Medical Informatics, 188, 105466. https://doi.org/10.1016/j.ijmedinf.2024.105466 (14.05.2025).

Kohli, A. (2020). “Snowballing” in Systematic Literature Review | LinkedIn. (n.d.). https://www.linkedin.com/pulse/snowballing-systematic-literature-review-amanpreet-kohli/ (14.05.2025).

Künnapuu, K., Ioannou, S., Ligi, K., Kolde, R., Laur, S., Vilo, J., Rijnbeek, P. R., & Reisberg, S. (2022). Trajectories: A framework for detecting temporal clinical event sequences from health data standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model. JAMIA Open, 5(1), ooac021. https://doi.org/10.1093/jamiaopen/ooac021 (10.02.2025).

Lin, V., Tsouchnika, A., Allakhverdiiev, E., Rosen, A. W., Gögenur, M., Clausen, J. S. R., Bräuner, K. B., Walbech, J. S., Rijnbeek, P., Drakos, I., & Gögenur, I. (2022). Training prediction models for individual risk assessment of postoperative complications after surgery for colorectal cancer. Techniques in Coloproctology, 26(8), 665–675. https://doi.org/10.1007/s10151-022-02624-x (31.03.2025).

Mellor, L. (2022). Cochrane Handbook for Systematic Reviews of Interventions.

Mourão, E., Pimentel, J. F., Murta, L., Kalinowski, M., Mendes, E., Wohlin, C. (2020). On the performance of hybrid search strategies for systematic literature in software engineering. https://doi.org/10.1016/j.infsof.2020.106294 (17.01.2025).

Oja, M., Tamm, S., Mooses, K., Pajusalu, M., Talvik, H.-A., Ott, A., Laht, M., Malk, M., Lõo, M., Holm, J., Haug, M., Šuvalov, H., Särg, D., Vilo, J., Laur, S., Kolde, R., & Reisberg, S. (2023). Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: Lessons learned. JAMIA Open, 6(4), ooad100. https://doi.org/10.1093/jamiaopen/ooad100 (14.05.2025).

Piao, X., Gao, P., Chen, Z. (2024). BEHRNOULLI: A Binary EHR Data Oriented Medication Recommendation System. (n.d.). Retrieved May 14, 2025, from https://arxiv.org/html/2408.09410v1 (14.05.2025).

Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. IEEE ASSP Magazine, 3(1), 4–16. https://doi.org/10.1109/MASSP.1986.1165342 (31.03.2025).

Reps, J. M., Schuemie, M. J., Suchard, M. A., Ryan, P. B., & Rijnbeek, P. R. (2018). Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. Journal of the American Medical Informatics Association, 25(8), 969–975. https://doi.org/10.1093/jamia/ocy032 (10.02.2025).

Romeu, J. L., (2020). A Markov Chain Model for Covid-19 Survival Analysis. https://doi.org/10.13140/RG.2.2.36349.18408 (17.03.2025).

Siggaard, T., Reguant, R., Jørgensen, I. F., Haue, A. D., Lademann, M., Aguayo-Orozco, A., Hjaltelin, J. X., Jensen, A. B., Banasik, K., & Brunak, S. (2020). Disease trajectory browser for exploring temporal, population-wide disease progression patterns in 7.2 million Danish patients. Nature Communications, 11(1), 4952. https://doi.org/10.1038/s41467-020-18682-4 (17.02.2025).

Soper, B. C., Nygård, M., Abdulla, G., Meng, R., & Nygård, J. F. (2020). A hidden Markov model for population-level cervical cancer screening data. Statistics in Medicine, 39(25), 3569–3590. https://doi.org/10.1002/sim.8681 (17.03.2025).

Tatman, R., VanderPlas, J., & Dane, S. (s.a.). A Practical Taxonomy of Reproducibility for Machine Learning Research.

Uhry, Z., Hédelin, G., Colonna, M., Asselain, B., Arveux, P., Rogel, A., Exbrayat, C., Guldenfels, C., Courtial, I., Soler-Michel, P., Molinié, F., Eilstein, D., & Duffy, S.

(2010). Multi-state Markov models in cancer screening evaluation: A brief review and case study. Statistical Methods in Medical Research, 19(5), 463–486. https://doi.org/10.1177/0962280209359848 (17.03.2025).

Understanding OMOP Basics. (2024, november 26). User Support. https://support.researchallofus.org/hc/en-us/articles/360039585391-Understanding-OMOP-Basics (07.04.2025).

Viscondi, J. Y. K., Faustino, C. G., Campolina, A. G., Itria, A., & Soárez, P. C. de. (2018). Simple but not simpler: A systematic review of Markov models for economic evaluation of cervical cancer screening. Clinics (Sao Paulo, Brazil), 73, e385. https://doi.org/10.6061/clinics/2018/e385 (13.04.2025).

Wang, S. V., Sreedhara, S. K., & Schneeweiss, S. (2022). Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. Nature Communications, 13(1), 5126. https://doi.org/10.1038/s41467-022-32310-3 (14.05.2025).

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLOS ONE, 12(4), e0174944. https://doi.org/10.1371/journal.pone.0174944 (10.02.2025).

Whitt, W. (2014). CONTINUOUS-TIME MARKOV CHAINS. https://yunanliu.wordpress.ncsu.edu/files/2018/07/CTMCnotes120614.pdf (31.03.2025).

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, 1–10. https://doi.org/10.1145/2601248.2601268 (14.05.2025).

Wohlin, C., Kalinowski, M., Felizardo, K. R., & Mendes, E. (2022). Successful combination of database search and snowballing for identification of primary studies in systematic literature studies. Information and Software Technology, 147, 106898. https://www.sciencedirect.com/science/article/pii/S0950584922000659#bib1 (5.12.2024).

Yang, C., Williams, R. D., Swerdel, J. N., Almeida, J. R., Brouwer, E. S., Burn, E., Carmona, L., Chatzidionysiou, K., Duarte-Salles, T., Fakhouri, W., Hottgenroth, A., Jani, M., Kolde, R., Kors, J. A., Kullamaa, L., Lane, J., Marinier, K., Michel, A., Stewart, H. M., … Rijnbeek, P. R. (2022). Development and external validation of prediction models for adverse health outcomes in rheumatoid arthritis: A multinational real-world cohort analysis. Seminars in Arthritis and Rheumatism, 56, 152050. https://doi.org/10.1016/j.semarthrit.2022.152050 (31.03.2025).

Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Kirke Valt**,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

   Andmepõhiste ravitrajektooride konstrueerimise meetodite süstemaatiline analüüs,

   mille juhendaja on **Markus Haug**,

   reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada Tartu Ülikooli digitaalarhiivi kuni autoriõiguse kehtivuse lõppemiseni;

2. annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi kaudu Creative Commonsi litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni;

3. olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile;

4. kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Kirke Valt

15.05.2025