

TARTU ÜLIKOOL
Loodus- ja täppisteaduste valdkond
Arvutiteaduse instituut
Andmeteaduse õppekava

Markus Haug

Patsientide ravitrajektooride modelleerimine Markovi ahelatega

Magistritöö (15 EAP)

Juhendaja: Raivo Kolde, PhD

Tartu 2022

Patsientide ravitrajektooride modelleerimine Markovi ahelatega

Lühikokkuvõte:

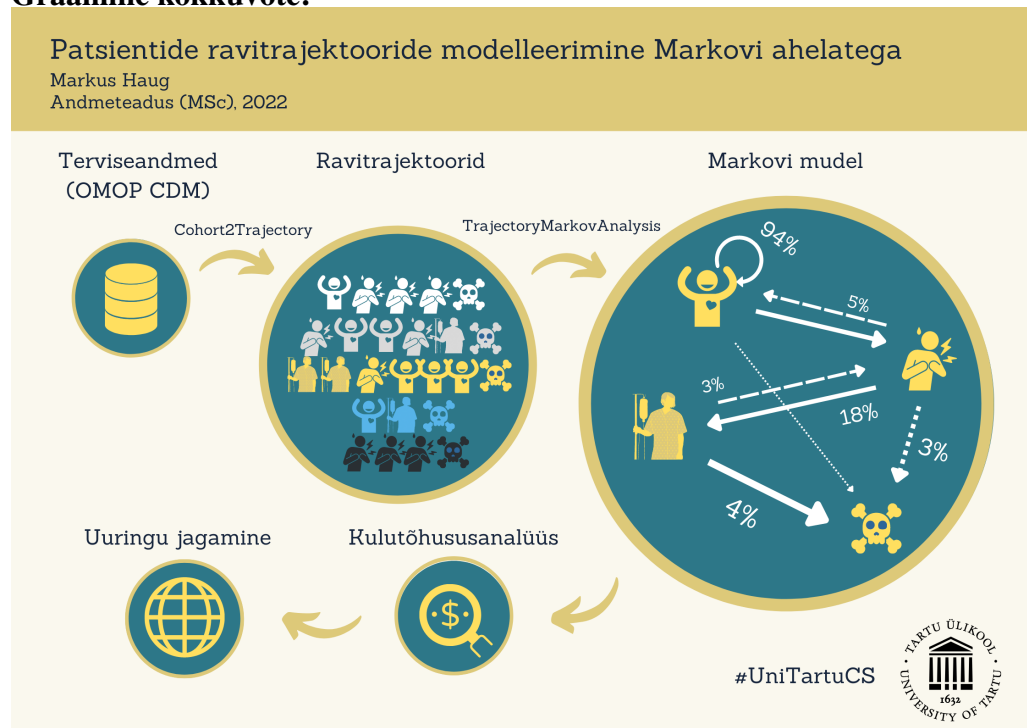
Suur hulk digitaliseeritud hoiustatud meditsiiniliseid andmeid tekitab võimaluse patsientide ravitrajektooride uurimiseks. Trajektooride uurimine on aluseks erinevate tervishoiuülesannete lahendamiseks, nagu näiteks prima tervishoiupraktika selgitamiseks, tervishoiuökonomika hindamiseks ja ravitrajektooride modelleerimiseks. Käesoleva magistritöö eesmärgiks on eelmainitud ülesannete lahendamiseks luua universaalsed tööriistad. Töö käigus arendatakse välja kaks R-paketti. Esimene pakett tekitab patsientide ravitrajektoore. Teine pakett põhineb trajektooride modelleerimisel Markovi ahelatega. Valminud tööriistad on kasutatavad kõikidel andmebaasisüsteemidel, mis sisaldavad endas vabavaralist *Observational Health Data Sciences and Informatics Observational Medical Outcomes Partnership Common Data Model* (edaspidi OHDSI OMOP CDM) andmemudelit. Töö käigus demonstreeritakse loodud pakettide kasutamist südamepuudulikkusega patsientide andmetel, mis pärinevad Eesti Haigekassast.

Võtmesõnad:

terviseandmed, ravitrajektoorid, Markovi ahel, OMOP CDM, südamepuudulikkus

CERCS: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlusmatemaatika.

Graafiline kokkuvõte:



Patient Treatment Trajectory Modeling With Markov Chains

Abstract:

The vast amount of electronically stored medical data offers a possibility to investigate patients' treatment trajectories. These trajectories give us a foundation to find out the best healthcare practices, evaluate the economics of treatment patterns and model the treatment paths. The aim of this work is to implement a technical solution for the aforementioned tasks. In the course of this thesis two R-packages will be developed. The first package uses electronic health data to construct treatment trajectories. The second package implements these trajectories and composes Markov models. The achieved tools are distributable to any database system operating on the open source Observational Health Data Sciences and Informatics Observational Medical Outcomes Partnership Common Data Model. Our universal solutions will be demonstrated on heart failure data supplied by Estonian Health Insurance Fund.

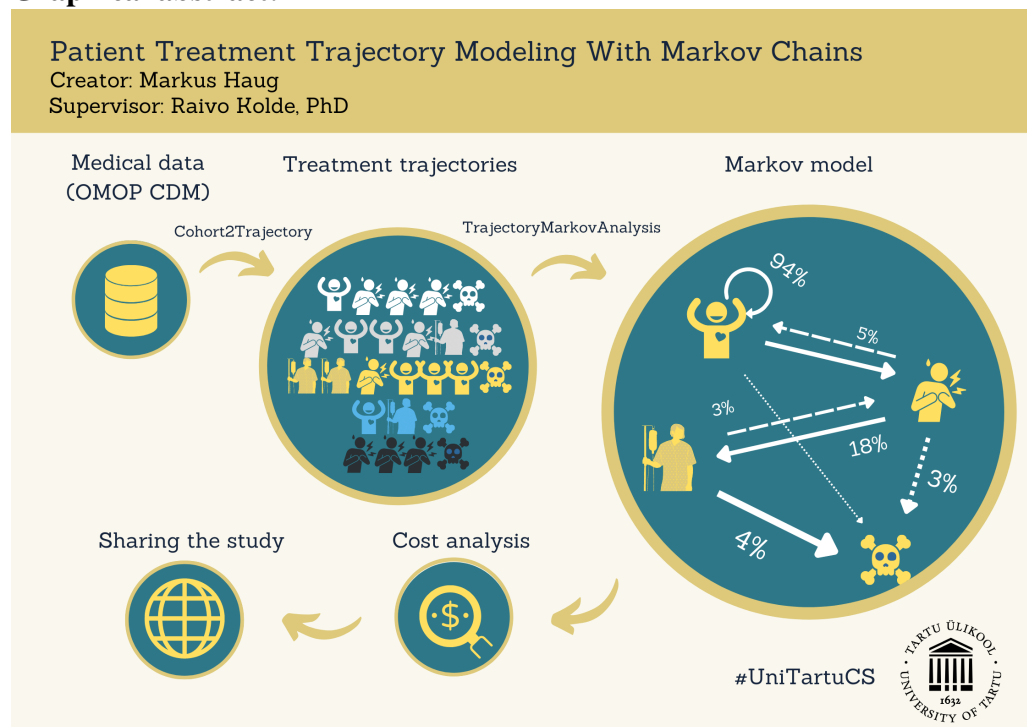
Keywords:

medical data, treatment trajectories, Markov chain, OMOP CDM, heart failure

CERCS:

P160 Statistics, operations research, programming, financial and actuarial mathematics.

Graphical abstract:



Sisukord

1	Sissejuhatus	6
2	Motivatsioon	8
3	Metoodika	10
3.1	Markovi ahelad	10
3.1.1	Diskreetse ajaga protsessid	11
3.1.2	Pideva ajaga protsessid	12
3.1.3	Markovi ahela seisunditest	13
3.2	Terviseökonomika	14
3.2.1	Analüüsid Markovi ahelatel	15
3.3	OHDSI OMOP ühtne andmemudel	18
3.3.1	OHDSI	18
3.3.2	OMOP ühtne andmemudel	18
3.3.3	ATLAS	19
3.3.4	R-paketid OHDSI kogukonnas	20
4	Implementatsioon	22
4.1	R-pakett <i>Cohort2Trajectory</i>	22
4.1.1	Paketi sisu	23
4.1.2	Trajektooride valmistamine	25
4.1.3	Võimalikud arendused	28
4.2	R-pakett <i>TrajectoryMarkovAnalysis</i>	29
4.2.1	Paketi sisu	29
4.2.2	Võimalikud arendused	32
5	Näidisuuring: südamepuudulikkusega patsientide ravitrajektoorid	34
5.1	Andmed	34
5.2	Tulemused	36
6	Kokkuvõte	41
	Viidatud kirjandus	42
	Lisad	46
1	Paketi <i>Cohort2Trajectory</i> kasutajaliidese vaheleht "Importimine ATLAS'st"	46
2	Paketi <i>Cohort2Trajectory</i> kasutajaliidese vaheleht "Importimine JSON'st"	47

3	Paketi <i>Cohort2Trajectory</i> kasutajaliidese vaheleht "Statistika"	48
4	Paketi <i>Cohort2Trajectory</i> kasutajaliidese vaheleht "Prioritiseerimine"	49
5	Paketi <i>Cohort2Trajectory</i> kasutajaliidese vaheleht "Trajektoorid"	50
6	Paketi <i>Cohort2Trajectory</i> kasutamine: ATLAS-tarkvaras kohordi defineerimine	51
7	Paketi <i>Cohort2Trajectory</i> kasutamine: näide väljastatud trajektooride failist	54
8	Paketi <i>TrajectoryMarkovAnalysis</i> kasutajaliidese vaheleht "Importimine"	55
9	Paketi <i>TrajectoryMarkovAnalysis</i> kasutajaliidese vaheleht "Diskreetse ajaga Markovi ahelad"	56
10	Paketi <i>TrajectoryMarkovAnalysis</i> kasutajaliidese vaheleht "Pideva ajaga Markovi ahelad"	59
11	Paketi <i>TrajectoryMarkovAnalysis</i> kasutajaliidese vaheleht "Kaplan-Meieri graafikud"	62
12	Paketi <i>TrajectoryMarkovAnalysis</i> kasutajaliidese vaheleht "Otsustuspuud"	63
13	Paketi <i>TrajectoryMarkovAnalysis</i> kasutajaliidese vaheleht "Profiilid"	64
14	Paketi <i>TrajectoryMarkovAnalysis</i> kasutajaliidese vaheleht "Genereerimine"	65
15	Litsents	67

1 Sissejuhatus

Terviseandmete digitaliseerimine on endaga kaasa toonud ohtralt võimalusi personaalmeditsiini edendamiseks. Digitaliseerimise tulemusena on tekkinud mitmeid kliiniliste otsuste tugisüsteeme ning on lihtsustunud meditsiinilise informatsiooni jagamine. Tänu sellele on parandatud patsientide ravikvaliteeti, vähendatud haigestumist ja redutseeritud meditsiinilisi ebatäpsusi [Menachemi ja Collum, 2011].

Kuigi elektrooniliste terviseandmete kasutuselevõtmisega on kergendatud andmete jagamist erinevate andmehoidlate vahel, on siiski raskusi erinevate andmestandardite ühendamisel. Üheks sellise probleemiga tegeleva rahvusvahelise kogukonnana on esile kerkinud OHDSI (*Observational Health Data Sciences and Informatics*). OHDSI on võtnud oma eesmärgiks luua ja rakendada tänapäevasele suurele terviseandmebaaside võrgustikule avatud lähtekoodiga andmeanalüütilisi lahendusi. Nende lahenduste abil edendatakse andmehoiustajate, teadlaste, tervishoiutöötajate ja patsientide võimalusi, mille resultaadiks on näiteks ühtne tervishoiuandmete mudel, paremini reprodutseeritav teadus, erinevad meditsiiniliste otsuste tugisüsteemid, personaalmeditsiini paranemine ja palju muud [Hripcsak et al., 2015].

Personaalmeditsiini üheks osaks on patsientide ravitrajektooride uurimine. Patsiendi ravitrajektoor koosneb ajaliselt järjestatud põetud haigustest, kasutatud ravimitest ja teenustest, tehtud visiitidest ja muudest meditsiinilistest sekkumistest. Trajektooride uurimine on olnud edukas näiteks nii haigustrajektooride modelleerimisel [Soper et al., 2020] kui tervishoiuökonoomika uurimisel [Kobelt, 2013].

Meditsiiniliste andmete modelleerimisel kasutatakse tihti Markovi mudeleid. Markovi ahelaid rakendatakse muuhulgas nii kliinilise riskiga sündmuste prognoosimiseks, alternatiivsete ravimeetodite efektiivsuse ja kulutõhususe võrdlemiseks. Mudel eeldab, et eksisteerib lõplik arv võimalikke seisundeid ning patsient on alati ühes kindlas seisundis. Selline lähenemine sobib hästi patsientide ravitrajektooride uurimiseks, kuna igasuguseid meditsiinilisi sekkumisi saab kirjeldada kui eraldiseisvaid seisundeid. Markovi mudeleid on edukalt implementeeritud meditsiinilistel andmetel nii COVID-19 uuringutel [Romeu, 2020], emakakaelavähki haigestumise modelleerimiseks [Soper et al., 2020], ravimite kontsentratsiooni optimeerimiseks [Ma et al., 2021] kui südamepuudulikkuse häirega haigestunute analüüsimiseks [Thokala et al., 2020].

Käesoleva rakendusliku magistritöö tulemusena valmib kaks R-paketti, mis töötavad OHDSI loodud lahendustel. Esimene tööriist keskendub patsientide ravitrajektooride moodustamisele. Teine tööriist kasutab esimese väljundit ning tekitab trajektooride põhjal Markovi mudeleid. Pakettide tööd ilmestatakse näidisuuringuga südamepuudulikkusega

patsientide andmestikul.

Magistritöö koosneb kuuest peatükist. Esimeses ja teises peatükis anname ülevaate teema aktuaalsusest ning motivatsioonist. Kolmandas peatükis avame tausta töös kasutatud meetodikatele: anname lihtsustatud ülevaate Markovi ahelate teooriast, kirjeldame terveökonoomika analüüsi läbiviimist meditsiinilistel andmetel ja tutvustame lugejale OMOP ühtset andmemudelit. Neljandas peatükis kirjeldame kahe töö käigus loodud R-paketi (*Cohort2Trajectory* ja *TrajectoryMarkovAnalysis*) funktsionaalsuseid. Viiendas peatükis demonstreerime eelmainitud pakettide tööd näidisuringu abil. Viimases, kuuendas, peatükis võtame käesoleva magistritöö kokku.

Magistritöö läbiviimiseks on kasutatud statistikatarkvara R (versioon 4.0.3). Töö on kirjutatud kasutades LaTeX veebirakenduse liidest Overleaf. Töös esinevad joonised on koostatud kasutades kujundusgraafika tarkvara Canva. Töö ilmestamiseks esitame näiteid Suurbritannia andmestikul läbiviidud uuringust, kus modelleeritakse Markovi ahelatega südamepuudulikkusega patsientide ravitrajektoore [Thokala et al., 2020].

2 Motivatsioon

Tervishoiu vaatlusuuringute andmed koosnevad digilugudest, tervisekindlustuse nõuetest, retseptidest ning muudest võimalikest meditsiinilistest ülestähendustest. Kuigi need andmed on tihti erinevatel standarditel ja mittetäielikud, on need teaduse perspektiivis muutumas aina tähtsamaks. Selleks, et andmetega teaduse tegemine oleks lihtsam, on Euroopa Komisjon valmistamas ette Euroopa tervishoiu andmeruumi regulatsioone, mille eesmärgiks on luua ühtne tugev andmehaldussüsteem, tagada andmete kvaliteet ning taristu ja koostalitlusvõime [European Commission, 2022]. Sarnaselt on Euroopa Raviamet algatanud mitmeid projekte vaatlusuuringute andmete kasutamiseks oma regulatsioonide määramisel.

Üha suurema huviga kasutatakse vaatlusandmeid, et hinnata meditsiiniliste tehnoloogiate turvalisust, efektiivsust ja kulutõhusust. Kahjuks piiravad operatiivsed, tehnilised ja meetodilised väljakutsed vaatlusandmete kasutamise laiemat levikut. On leitud, et potentsiaalseks lahenduseks võiksid olla ühisele standardile viidud andmemudelid. Avatud lähtekoodiga OMOP ühtne andmemudel standardiseerib varasemalt ühendamatuid andmebaaside struktuure, formaate ja terminoloogiasid. See andmemudel on suuteline hõlbustama juurdepääsu asjakohastele andmetele, võimaldama mitme andmebaasiga uuringuid ja valideerima tulemusi. Avatud lähtekoodi ja standardiseeritud analüütika tööriistade kasutamine parandab otsustusprotsesside läbipaistvust ja suurendab publitseeritud tulemuste usaldusväärsust [Kent et al., 2021]. Tänu OMOP ühtsetele andmeformaatile on teoks saanud mitmed erinevad metodoloogilised uuringud nagu teadustööd juhtkontrolluuringutest [Schuemie et al., 2019], allikate integreerimisest [Voss et al., 2017], kontrollide sobitamisest [Tian et al., 2018] ja paljust muust. Standardiseeritud tarkvara loomisega on optimeeritud teadustöö tegemist populatsiooni- ja kohordipõhiste andmetega [Hripcsak et al., 2021].

Senimaani on OHDSI keskendunud populatsiooni- ja patsiendipõhiste tulemuste kirjeldamisele. Vähem tööd on tehtud ravitrajektooride kirjeldamise ja kvantifitseerimisega. Ravitrajektoori kirjeldavad patsiendiga toimunud kliiniliste sekkumiste jadad. Selle osaks võivad olla sümptomid, mõõtmised, diagnoosid, raviprotseduurid, ravimite manustamine või muu selline. Individuaalselt on sellised sündmused tüüpiliselt seotud ajaga, sellest ka analoogia jadadega. Hetkel teeb selliste jadade tuntud mudelitega uurimise raskeks erinevad seisundite kestused, nendevaheline aeg, ebaselge agregeerimine ja puuduvad andmed.

Meditiinilisest vaatepunktist trajektooride uurimine on huvitav mitmest aspektist. Esiteks pakuvad peale andmete kvantitatiivse analüüsi sarnased ravitrajektooriid head ülevaadet andmetest [Künnapuu et al., 2022]. Teiseks võimaldab trajektooride analüü-

simine leida tõendeid populatsioonis ravi rakendamise ja selle tulemuslikkuse kohta [Hripcsak et al., 2016]. Kliiniliste teekondade uurimisel saab leida ühiseid riskifaktoreid [Hu et al., 2019] või haiguste lähtekohti [Giannoula et al., 2020]. Selliste väärtuslike teadmiste leidmisel saaks parandada kliinilist ravi või teisi meditsiinilisi strateegiaid. Viimaseks on trajektooride analüüsimine potentsiaalselt kasulik individuaalsete haigusjuhtude ennustamiseks ja ärahoidmiseks [Jensen et al., 2014].

Eelnevale lisaks on ravitrajektoore kasutades võimalik uurida ka tervishoiuökonoomikat. Uurides erinevate tervise seisundite esinemist ja maksumust, saab võrrelda erinevaid ravimeetodeid tehes kulutõhususe analüüsi. Selleks vajalike Markovi mudelite koostamiseks saab kasutada ravitrajektoories peituvat infot [Kobelt, 2013].

Senimaani on loodud OMOP ühtse andmemudelile mõningaid patsientide trajektoore uurivaid R-pakette. Näiteks pakett *TreatmentPatterns* [Markus, 2021], mille põhilisteks funktsioonideks on kirjeldada ja visualiseerida kohordis esinevaid teekondi, ja *Trajectories* [Künnapuu et al., 2021], mis leiab kohordis ülesse statistiliselt olulised seisundite paarid. Sellised paketid tegelevad enamasti ainult kvantitatiivse analüüsiga. Nendel pakettidel puudub potentsiaal ravitrajektoore eelnevalt tutvustatud ülesannete (ennustusmudelid, tervishoiuökonoomika jms) lahendamiseks kasutada. Teisalt on tekkinud vajadus patsientide ravitrajektooride tekitamiseks. Näiteks pakett *TrajectoryViz* [Pajusalu, 2022] vajab sisendiks just patsientide ravitrajektoore.

Selleks, et realiseerida OMOP ühtse andmemudeli potentsiaali, tuleb panna rõhku seda toetavate tööriistade loomisele. Hetkel loodud R-paketid katavad vaid piiratud osa kõikidest võimalikest OMOP andmemudelil rakendatavatest analüüsides. Puudu on tervishoiuanalüütikat võimaldavatest rakendustest, mis toetaksid näiteks kulude ja elukvaliteedi modelleerimist [Kent et al., 2021].

Eelnevast tulenevalt tundub oluline tekitada OMOP ühtse andmemudelile tööriistade süsteem, mille eesmärgiks on ravitrajektooride analüüsimine. Selleks oleks tarvilik luua töövoog, mis väljastaks kohortide põhjal sinna sattunud patsientide ravitrajektooreid. Lisaks sellele loome töövoole funktsionaalsused trajektooride modelleerimiseks, mille eesmärgiks oleks eelmainitud ravitrajektooride analüüsimise potentsiaali täitmine.

3 Metoodika

Käesolevas peatükis anname ülevaate Markovi ahelate teoriast, kirjeldame terviseökonoomika analüüsi läbiviimist meditsiinilistel andmetel ja tutvustame lugejale OMOP ühtset andmemudelit. Markovi ahelate teooria on aluseks patsientide ravitrajektooride Markovi mudelite koostamisel. Terviseökonoomika tausta avamine on tähtis ravikulutuste uurimiseks ja võrdlemiseks. OMOP ühtne andmemudel on terviseandmete hoiustamise ning antud magistritöö aluseks.

3.1 Markovi ahelad

Markovi ahelad on vene matemaatiku Andrei Markovi teoretiseeringute tulemus 20. sajandi algusest. Need on stohhastilised protsessid, mis kirjeldavad üleminekuid ühest olekust teise. Aastate jooksul on need leidnud lugematul hulgal rakendust protsesside modelleerimises ja informeeritud otsuste tegemisel. Markovi ahelaid kasutatakse füüsikas, järjekorrateoorias, rahanduses, sotsiaalteadustes, statistikas ja loomulikult meditsiinis. Markovi mudelid on kasulikud keskkondade ja probleemide modelleerimisel, mis hõlmavad aja jooksul järjestikuseid stohhastilisi liikumisi [Komorowski ja Raffa, 2016].

Järgnevalt anname lihtsustatud ülevaate Markovi ahelatest toetudes J. R. Norrise raamatule "Markov Chains" [Norris, 1998] ja Meelis Kääriku kursuse "Juhuslikud protsessid" loengukonspektile [Käärik, 2017].

Markovi protsessiks nimetatakse juhuslikku protsessi, mis on mäluta protsess ehk ajas järgmine seisund oleneb ainult hetkelisest seisundist. Markovi ahel on Markovi protsess, millel on lõplik või loenduv arv võimalikke seisundeid.

Markovi ahelaid saab koostada diskreetse ajaga

$$n \in \mathbb{Z}^+ = \{0, 1, 2, \dots\}$$

ja pideva ajaga

$$t \in \mathbb{R}^+ = [0, \infty).$$

Edaspidi märgistamegi valemiga $(X_n)_{n \geq 0}$ diskreetse ajaga protsessi ja valemiga $(X_t)_{t \geq 0}$ pideva ajaga protsessi.

3.1.1 Diskreetse ajaga protsessid

Olgu I loenduv hulk. Igat elementi $i \in I$ kutsume edaspidi seisundiks. Defineerime suuruse $\lambda = (\lambda_i : i \in I)$, mis kirjeldab hulka I , kui $0 \leq \lambda_i < \infty$ kõikide $i \in I$ puhul. Juhul, kui $\sum_{i \in I} \lambda_i = 1$, siis kutsume λ jaotuseks, kus

$$\lambda_i = \mathbb{P}(X = i).$$

Seega on X juhuslik seisund, mis võtab väärtuse i tõenäosusega λ_i . Maatriksit $P = (p_{ij} : i, j \in I)$ kutsutakse stohhastiliseks, kui iga rida $(p_{ij} : j \in I)$ on jaotus.

Defineerime nüüd Markovi ahela. Me ütleme, et $(X_n)_{n \geq 0}$ on Markovi ahel algse jaotusega λ ja üleminekumaatriksiga P , kui

1. X_0 on jaotusega λ ;
2. kui $n \geq 0$ ja $X_n = i$, siis X_{n+1} on jaotusega $(p_{ij} : j \in I)$ ja on sõltumatu seisunditest X_0, \dots, X_{n-1} .

Teiste sõnadega – kui $n \geq 0$ ja $i_1, \dots, i_{n+1} \in I$, siis

1. $\mathbb{P}(X_0 = i_1) = \lambda_{i_1}$;
2. $\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_1, \dots, X_n = i_n) = p_{i_n i_{n+1}}$.

Diskreetse ajaga juhuslik protsess $(X_n)_{0 \leq n \leq N}$ on Markovi ahel siis ja ainult siis, kui iga $i_1, \dots, i_N \in I$ korral kehtib

$$\mathbb{P}(X_0 = i_1, X_1 = i_2, \dots, X_N = i_N) = \lambda_{i_1} p_{i_1 i_2} p_{i_2 i_3} \dots p_{i_{N-1} i_N}.$$

Olgu meil homogeenne (diskreetse ajaga) Markovi ahel üleminekumaatriksiga $P = (p_{ij})$ ja olgu meie ülesandeks leida tõenäosus

$$P(\text{seisundist } i \text{ seisundisse } j \text{ } k \text{ sammu jooksul}) =: p_{ij}(k).$$

Tuleb välja, et diskreetse Markovi ahela puhul kehtib $p_{ij}(k) = P_{ij}^k$.

Tabel 1. Markovi ahela näide, kus tervise seisundid 0,1,2,3 ja 4 (surm) kirjeldavad haiguse progresseerumist

Markovi ahel					
Tervise seisund	0	1	2	3	4 (surm)
0	0,981	0,004	0,000	0,000	0,015
1	0,068	0,872	0,017	0,000	0,043
2	0,004	0,094	0,777	0,041	0,084
3	0,000	0,006	0,095	0,777	0,122
4	0,000	0,000	0,000	0,000	1,000

3.1.2 Pideva ajaga protsessid

Pideva ajaga protsessid on sarnased diskreetse ajaga protsessidele. Erinevus seisneb selles, et pideva ajaga protsesside puhul on igal Markovi seisundil erinev olekus viibimise aeg.

Olgu I loenduv hulk. Hulgale I loodud Q -maatriks, kus $Q = (q_{ij} : i, j \in I)$ rahuldab järgmisi tingimusi:

- $0 \leq q_{ii} < \infty, \forall i$;
- $q_{ij} \geq 0, \forall i \neq j$;
- $\sum_{j \in I} q_{ij} = 0, \forall i$.

Seega igal Q -maatriksi real

$$q_i = \sum_{j \neq i} q_{ij} \leq \infty,$$

kus $q_i = -q_{ii}$. Maatriksit Q kutsutakse intensiivsusmaatriksiks ning väärtust q_i seisundist i lahkumise intensiivsuseks, mis omakorda kirjeldab seisundis $i \in I$ olemise aega, mis on juhuslik suurus eksponentjaotusega $Exp(q_i)$.

Igale intensiivsusmaatriksile Q vastab stohhastiline maatriks P . Üleminekumaatriks $P = (p_{ij} : i, j \in I)$, mis põhineb Q 'l, väärtused on

$$p_{ij} = \begin{cases} q_{ij}/q_i & \text{kui } j \neq i \text{ ja } q_i \neq 0 \\ 0 & \text{kui } j \neq i \text{ ja } q_i = 0, \end{cases}$$

$$p_{ii} = \begin{cases} 0 & \text{kui } q_i \neq 0 \\ 1 & \text{kui } q_i = 0. \end{cases}$$

Pideva ajaga juhuslik protsess

$$(X_t)_{t \leq 0} = (X_t : 0 \leq t < \infty)$$

väärtustega hulgast I on juhuslike suuruste pere $X_t : \Omega \rightarrow I$. Seda kutsutakse pideva ajaga Markovi ahelaks algse jaotusega λ ja intensiivsusmaatriksiga Q , mis

1. sattudes seisundisse $i \in I$ jääb ahel sinna ajaks $Exp(q_i)$, mis on sõltumatu juhuslik suurus eksponentjaotusest.
2. väljudes seisundist i , siseneb ta teatava tõenäosusega P_{ij} seisundisse j , kusjuures $P_{ii} = 0$ ja $\sum_j P_{ij} = 1$ iga i korral.

Seega pideva ajaga Markovi ahel on juhuslik protsess, kus ühest olekust teisse liikumine toimub nagu diskreetse ajaga Markovi ahelas, kuid igas olekus viibimise aeg on eksponentjaotusega juhuslik suurus.

Tähistame tõenäosuse, et protsess, mis hetkel s viibib olekus i , on aja t möödudes olekus j , järgmiselt:

$$P_{ij}(t) = P\{X(t+s) = j | X(s) = i\}.$$

Tuleb välja, et suvaliste $s, t \geq 0$ korral kehtib

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t) \cdot P_{kj}(s).$$

3.1.3 Markovi ahela seisunditest

Markovi ahela seisunditele omistatakse sõltuvalt nende esinemisele erinevaid omadusi. Järgnevalt toome välja mõned Markovi ahela seisundite iseärasused, mida võib kohata terviseandmete modelleerimisel.

Seisundit i , mille korral $p_{ii} = 1$, nimetatakse neelavaks seisundiks. See tähendab, et juhul kui ahel jõuab antud seisundisse, siis sellest seisundist ta enam ei välju. Meditsiinilisest vaatepunktist on üheks universaalsemaks näiteks patsiendi surma seisund.

Seisundit i nimetatakse ebaoluliseks, kui leidub täisarv $t_0 > 0$ ja leidub seisund j nii, et $p_{ij}(t_0) > 0$, kuid $p_{ji}(t) = 0 \forall t = 1, 2, \dots$ puhul. See tähendab, et peale ebaolulisest seisundist i väljumist olekusse j ei saa sinna enam kunagi tagasi. Vastupidiselt neelava seisundi näitele on siin heaks näiteks sündimine (patsient sünnib vaid korra). Kõiki seisundeid, mis ei ole ebaolulised, nimetatakse olulisteks.

Markovi ahelate seisunditevahelisi üleminekutõenäosusi arvutatakse andmetel suurima tõepära meetodiga. Tuleb välja, et Markovi ahelate puhul taandub see valemiks

$$p_{ij} = \frac{n_{ij}}{n_i},$$

kus n_{ij} tähistab kõikide liikumiste arvu seisundist i seisundisse j ja n_i seisundi i esinemus.

Selleks, et leida seisundite jaotus n tsükli jooksul seisundite algjaotusest v_0 , saab kasutada tõenäosusmaatriksit P . Tähistame seisundite jaotuse n tsükli hetkel v_n . Algjaotuse v_0 puhul kehtib

$$v_n = v_0 P^n = v_1 P^{n-1} = \dots = v_{n-1} P. \quad (1)$$

Kasutades Markovi mudeli algjaotust ja üleminekumaatriksit saab tekitada sünteetilisi trajektoore. Markovi ahelate Monte Carlo meetodeid rakendades on võimalik tekitada juhuslikke seisundite järjestusi, mis vastavad Markovi ahela poolt kirjeldatavatele protsessidele. Selliseid lähenemisi praktiseerivad näiteks epidemioloogid [Hamra et al., 2013]

3.2 Terviseökonomika

Järgnevalt annan ülevaate terviseökonomikast toetudes Gisela Kobelti väljaandele "Health Economics: An Introduction to Economic Evaluation" [Kobelt, 2013].

Terviseökonomika üheks eesmärgiks on hinnata raha jaotumist ja jaotamist tervishoius. Sellise majandusliku hindamise algpunktiks tervishoius on võrdlev analüüs alternatiivsete ravitrajektooride vahel nii hinna kui mõju põhjal. Näiteks saab sel moel võrrelda farmaatsiatooteid iseendaga (erinevate manustamiskoguste osas), konkureerivaid ravimeid omavahel või erinevat tüüpi sekkumisi patsiendi ravis. Selliseid analüüse tehes tuleb lisaks hinnale võtta arvesse nii patsiendi paranemise kiirust kui mõju elukvaliteedile.

Kõikvõimalikud majanduslikud analüüsid hõlmavad endas nii sisendi (ressursside kasutamine ja kaotus) kui väljundi hindamist (kasu tervisele). Terviseökonomika eesmärgiks on valida sobilikum alternatiiv, millest oleks kasu kogu mõjualas olevale populatsioonile. Tüüpiliselt on kõige ressursinõudlikum variant ka kõige suurema kasumääraga. Kuna ressursse ei jagu kõigile, tuleb hinnata, kas ravist saadav kasu tervisele on väärt suuremat investeeringut.

Olukorras, kus erinevad kliinilised sekkumised ei anna samu tulemusi, on tarvis hinnata võimalikke alternatiive nii kulude kui tagajärgede vaatepunktist. Üheks võimalikuks

viisiks erinevaid ravivorme majanduslikult võrrelda on kulutõhususe analüüsi (ingl *cost-effectiveness analysis*) tegemine. Selle käigus võrreldakse kulusid näiteks päästetud elude arvuga, eeldatavate eluaastate suurenemisega või elukvaliteedi paranemisega. Üheks analüüsivormiks on täiendkulu tõhususe määra (TTM; ingl *incremental cost-effectiveness ratio*) leidmine. Seda leitakse järgmise valemiga:

$$TTM = \frac{\text{kulud uue tehnoloogiaga} - \text{kulud vana tehnoloogiaga}}{\text{efekt uue tehnoloogiaga} - \text{efekt vana tehnoloogiaga}}$$

Toome näite südamepuudulikkuse uuringust [Thokala et al., 2020]. Raviks on ettenähtud täiendkulu 20000 £ ühe QALY (ingl *quality-adjusted life year*) ehk kvaliteetselt elatud eluaasta kohta. Meile on (kirjandusest või andmetest) teada, et telemeetria (uus tehnoloogia) kasutuselevõtt annab võrreldes tavalise raviga patsiendile juurde 0,075 QALY ja selle tehnoloogia rakendamine maksab 3610 £. Seega $TTM = \frac{3610}{0,075} \approx 48133 \text{ £/QALY}$, millest järeldame, et antud ravi ei saa võimaldatud täiendkulu raames kasutusele võtta.

Kobelt rõhutab, et TTM kasutamine ei pruugi olla sobilik praktikas kasutamiseks. Ühe põhjusena toob ta välja üleüldise andmete vähesuse, kuid leiab, et mitmete andmestike ühendamine sellisteks uuringuteks on üldiselt aksepteeritav.

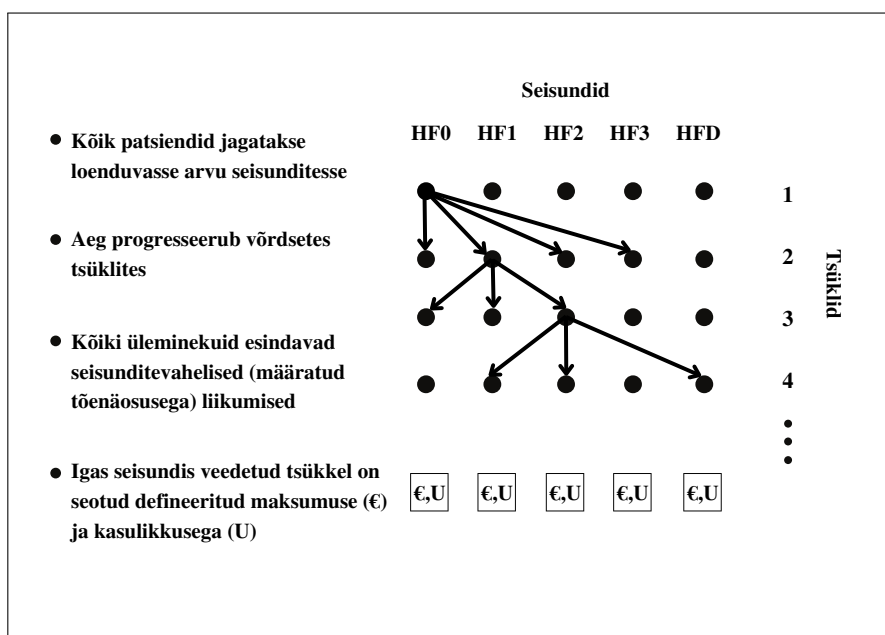
Ökonoomika hindamiseks kasutatakse tavaliselt järgmisi modelleerimise tüüpe:

1. Otsustuspuid kasutatakse kulutõhususe analüüsiks haiguste puhul, mille sündmused esinevad suhteliselt lühikese aja jooksul kindla tõenäosusega.
2. Markovi mudeleid kasutatakse analüüsiks juhtudel, kui haigus on läbiv pikemal perioodil.

3.2.1 Analüüsid Markovi ahelatel

Markovi mudelite puhul eeldatakse, et kõikidele patsientidele saab omistada loenduva hulga omavahel eristatavaid seisundeid. Tüüpiliselt defineeritakse need seisundid haiguse parameetrite (näiteks ägeduse taseme) järgi. Haiguse arengut või ravi efektiivsust kujutatakse seisunditevahelise liikumisena. Kui haiguse progressiooni iseloomustab liikumine kõrgemasse tasemesse, siis ravi toimimist liikumine samasse või madalamasse seisundisse. Modelleerimise jaoks on oluline, et seisundite eristus esindab kliinilisi erinevusi.

Markovi mudeli puhul on oluline ka see, et tsüklite pikkus peegeldab meditsiinilist loomulikku ajaintervalli. Näiteks – kui kasutada nädala pikkuseid tsikleid kõrgvererõhu



Joonis 1. Illustratsioon Markovi ahela analüüsist

ravi uurimiseks, oleks see liiga lühike periood. Samamoodi ei ole mõistlik uurida põletiku ravi aasta pikkuse intervalliga. Iga tsükli jooksul võib patsient liikuda ühest seisundist mõnda teise või jääda samasse seisundisse. Liikumiste tõenäosused arvutatakse tüüpiliselt epidemioloogilistest andmetest või leitakse vastavatest uuringutest.

Järgnevalt toome taas näite südamepuudulikkusega patsientide uuringust. Määrame tsükli pikkuseks ühe aasta. Olgu meil defineeritud viis kliinilist seisundit ning nende kulu tervishoiusüsteemile ja väärtus QALY'des:

- HF0 – patsiendil pole viimase aasta jooksul olnud ühtegi südamepuudulikkusega seotud hospitaliseerimist (300 €, 1 QALY),
- HF1 – patsiendil on viimase aasta jooksul olnud üks südamepuudulikkusega seotud hospitaliseerimine (4300 €, 0,75 QALY),
- HF2 – patsiendil on viimase aasta jooksul olnud kaks südamepuudulikkusega seotud hospitaliseerimist (8300 €, 0,5 QALY),
- HF3 – patsiendil on viimase aasta jooksul olnud kolm või enam südamepuudulikkusega seotud hospitaliseerimist (15300 €, 0,25 QALY),
- HFD – patsient suri (0 €, 0 QALY).

Tabel 2. Patsientide jaotus viie aasta jooksul tabelis 1 toodud Markovi mudeli järgi (ümarstatud kolme komakohani)

Patsientide jaotus viie aasta lõikes							
Tervise seisund Aasta (tsükkel)	HF0	HF1	HF2	HF3	HFD	Kum. QALY	Kum. ku- lud (€)
0	1	0	0	0	0	1	300
1	0,981	0,004	0	0	0,015	1,984	611,5
2	0,945	0,001	0	0	0,045	2,937	938
3	0,895	0,016	0,001	0	0,088	3,844	1283,6
4	0,834	0,021	0,001	0	0,144	4,694	1632,4
5	0,765	0,023	0,002	0	0,21	5,478	1977,4

Tabel 3. Patsientide jaotus viie aasta jooksul kasutades telemonitooringut

Telemonitooritud patsientide jaotus viie aasta lõikes							
Tervise seisund Aasta (tsükkel)	HF0	HF1	HF2	HF3	HFD	Kum. QALY	Kum. ku- lud (€)
0	1	0	0	0	0	1	450
1	0,986	0,003	0	0	0,011	1,988	908,7
2	0,959	0,008	0	0	0,033	2,953	1380,8
3	0,921	0,013	0	0	0,065	3,884	1863
4	0,874	0,017	0,001	0	0,107	4,771	2356,6
5	0,821	0,02	0,001	0	0,157	5,608	2847,2

Eeldame, et kõik patsiendid alustavad seisundist HF0. Kasutame tabelis 1 esitatud Markovi mudelit ning arvutame selle järgi patsientide jaotuse erinevates seisundites viie aasta lõikes (vt tabel 2). Joonisel 1 on kujutatud võimalikke liikumisi antud Markovi ahelas.

Oletame nüüd, et südamepuudulikkusega patsientide ravi saab muuta efektiivsemaks telemonitooringuga. Olgu, et telemonitooringu kasutamine maksab aastas 150 €. Vastava alternatiivi kasutamisel nähtav jaotus on toodud tabelis 3.

Tabelitest 2 ja 3 on näha, et alternatiivne ravi hoiab ära haiguse progresseerumist, kuid maksab märgatavalt rohkem. Arvutame täiendkulu tõhususe määra

$$TMM = \frac{2847,2 - 1977,4}{5,608 - 5,478} \approx 6690,77 \text{ €/QALY.}$$

Siit järeldub, et iga lisaks investeeritud 6690,77 € eest võidaksime me ühe kvaliteetse eluaasta. Juhul, kui telemeetría kasutamine ei hõlmaks endas lisakulutusi, siis $TMM \approx$

–232,31€/QALY, mis tähendaks, et alternatiivne ravi domineerib olemasoleva ravi üle (võidame nii rahas kui patsientide tervises).

3.3 OHDSI OMOP ühtne andmemudel

Järgnevalt anname ülevaate OHDSI OMOP ühtsest andmemudelist, millel põhineb suur osa käesolevast tööst. Antud peatükis kirjeldame lihtsustatult raamatus "The Book of OHDSI: Observational Health Data Sciences and Informatics" [OHDSI, 2021] leiduvat infot.

3.3.1 OHDSI

Observational Health Data Sciences and Informatics ehk OHDSI on 2014. aastal loodud avatud teaduse kogukond, mille eesmärk on parandada inimeste tervist, võimaldades kogukonnas ühtset tõendite kogumist, mis soodustavad paremate tervishoiuteenuste pakumist. OHDSI viib läbi metodoloogilisi uuringuid, mille eesmärk on luua teaduslikke parimaid tavasid vaatluslike terviseandmete kasutamiseks. Lisaks arendab kogukond avatud lähtekoodiga analüütikatarkvara, mis kodifitseerib eelmainitud tavad järjepidevateks, läbipaistvateks ja kergelt reprodutseeritavateks lahendusteks. Lahendusi rakendatakse kliinilistes küsimustes, et luua tõendeid, mis suunaksid tervishoiupoliitikat ja patsiendihooldust. Tänapäevaks on OHDSI kogukonnas üle 2500 aktiivse koostööpartneri rohkem kui 20 riigis. OMOP projekti poolt loodud ja nüüdse OHDSI keskmeks olevale OMOP ühtse andmemudeli formaati on viidud üle miljardi patsiendi sissekanded.

3.3.2 OMOP ühtne andmemudel

OMOP ühtne andmemudel loodi selleks, et lahendada tehnilisi probleeme teadustööde tegemisel, kus on kasutusel erineva struktuuriga vaatlusandmebaasid. Andmemudel on disainitud standardiseeritud vaatlusandmete struktuuri, sisu ja semantika kasutamiseks. Andmemudeli eesmärgiks on hõlbustada taaskasutatavate statistiliste analüüside tegemist kõikides patsiendipõhistes andmebaasides.

OMOP ühtse andmemudeli loomisest saati on kogukond arendanud seda avatud teaduse põhimõttel. Selleks, et edendada loodud tööriistade läbipaistvust ja kasvatada usaldust, on kogukonna kogu töö tulemused avalikud.

Sõltuvalt nende esmastest vajadustest ei kajasta ükski vaatlusandmebaas kõiki kliinilisi sündmusi võrdselt hästi. Seetõttu tuleb uurimistulemusi võtta paljudest erinevatest andmeallikatest, neid võrrelda ja vastandada, et avastada võimalikku kallutatuse mõju. Lisaks vajame statistilise võimsusega järelduste tegemiseks suurt hulka jälgitavaid patsiente. See selgitab vajadust hinnata ja analüüsida mitut andmeallikat samaaegselt. Selleks tuleb andmed viia samale andmestandardile. Lisaks nõuavad patsientide andmed kõrgetasemelist kaitset. Analüüsi eesmärgil andmete väljavõtmiseks, nagu seda traditsiooniliselt tehakse, on vaja rangeid andmekasutuslepinguid ja keerukat juurdepääsukontrolli. Ühine andmestandard võib seda vajadust leevendada, jättes välja ekstraheerimisetapi ja võimaldades teostada standardiseeritud analüütikat selle loomulikus keskkonnas olevate andmete põhjal. See tähendab, et andmeanalüütik saab ligipääsu andmebaasile, kus on juba analüüsimiseks korrastatud ja kaitstud andmed.

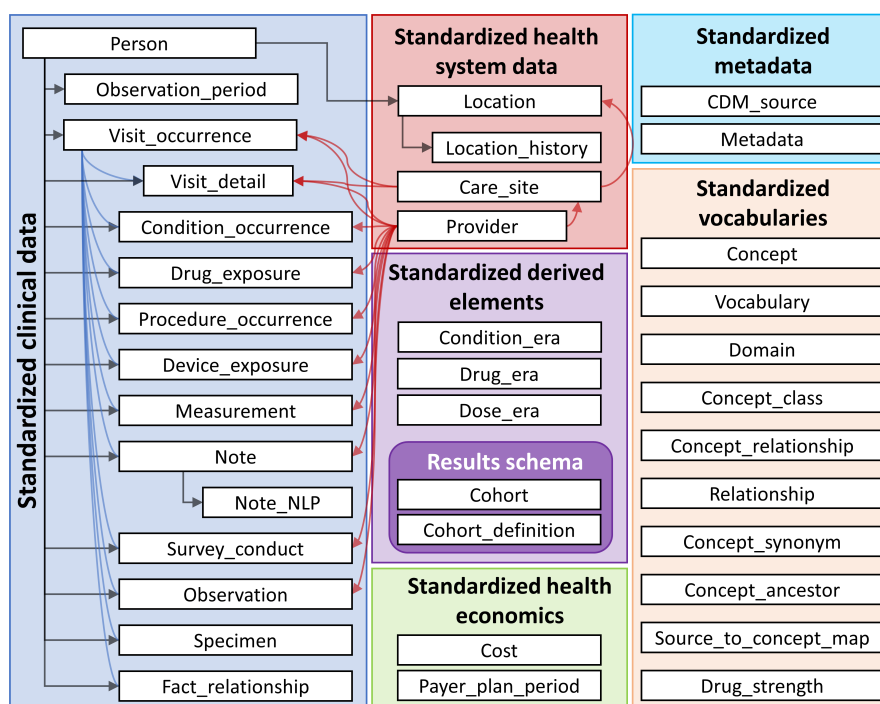
Selle standardi pakub OMOP ühtne andmemudel. Väljatoodud patsiendikeskne andmemudel mahutab endasse ära kõik asjakohased patsientidega seotud meditsiinilised sündmused. Andmemudel koos selle standardiseeritud sisuga tagab uurimismeetodite süstemaatilise rakendamise eesmärgiga saada võrreldavaid ja reprodutseeritavaid tulemusi.

Joonis 2 annab ülevaate kõikidest OMOP ühtsesse andmemudelisse kuuluvatest tabelitest. Nagu jooniselt näha, jagunevad tabelid omakorda erinevatesse skeemidesse. Antud töö seisukohast on kõige olulisem tabel *Cohort*, kust saab infot kohortide ja nendesse kuuluvate patsientide kohta. Tabelist *Cost* päritakse inimeste meditsiiniarvete maksumusi ning tabelist *Person* demograafilisi andmeid.

3.3.3 ATLAS

ATLAS on OHDSI loodud avatud lähtekoodiga tarkvaratööriist, millega saab viia läbi teaduslikke analüüse standardiseeritud vaatlusandmetega, mis on teisendatud OMOP ühtse andmemudeli kujule. Kasutajad saavad luua kohorte defineerides neid patsientide ravimiga kokkupuute, konkreetse haiguse diagnoosimise, arstivisiitide, jälgimisperioodi, surma, mõõtmiste, ravi eest tasumise või nende kombinatsioonide põhjal. ATLAS-rakenduses on otsingumootor, mis toetab meditsiiniliste mõistete otsimist, et tuvastada inimesi, kellel on teatud haigusseisundid, kokkupuude ravimitega ja palju muud. Patsiendiprofiile saab vaadata konkreetses rühmas, mis võimaldab visualiseerida konkreetse subjekti tervishoiuandmeid.

Antud töö kontekstis on tähtis ATLAS-rakenduse võimekus luua väga spetsiifilisi patsientide kohorte (vt lisa 6). Lisaks võimaldab ATLAS moodustatud kohorte kõikide OMOP ühtse andmemudeli kasutajatega jagada, väljastades loodud kohorte kirjeldava



Joonis 2. Ülevaade OMOP ühtse andmemudeli tabelitest [OHDSI, 2021]

SQL (*Structured Query Language*) või JSON (*JavaScript Object Notation*) kodeeringu. Eelmainitud funktsioone võib nimetada käesoleva magistritöö alusteks. Selline võimalus kohorte lihtsasti jagada ja võrrelda on reprodutseeritavale teadusele väga kasulik.

3.3.4 R-paketid OHDSI kogukonnas

OHDSI julgustab arendajaid ja muid huvilisi looma vaatlusandmetel põhinevaid uuringute pakette. Selle jaoks on loodud kollektsioon R-pakettidest, mis võimaldavad OMOP ühtsele andmemudelile luua kergesti reprodutseeritavaid ja läbipaistvaid teadustöid. Peamiselt jagunevadki OHDSI poolt hooldatavad paketid kaheks: töövoogu toetavad paketid ja uuringute paketid. Sellised OHDSI paketid on koondatud HADES-nimelisse raamatukokku, kus on eesmärgiks toetuda parimale teadaolevale vaatlusandmete ja kohortuuringute koostamise praktikale [OHDSI, 2021].

Töövoogu toetavateks pakettideks võib nimetada pakette, mis ise otseselt tulemusena teadustööd ei tee, kuid on näiteks reprodutseeritavuse tagamiseks ülimalt tähtsad. OHDSI vaatepunktist on sellised paketid näiteks *SqlRender* [Schuemie ja Suchard, 2022b]

ja *DatabaseConnector* [Schuemie ja Suchard, 2022a], mis tagavad võimaluse rakendada uuringute pakette erinevatel SQL dialektidel. Uuringute pakettideks võib nimetada pakette, mis toovad kasutajani vaatlusandmetest eraldatud teadmisi. Üheks selliseks OHDSI paketiks on *PatientLevelPrediction* [Reps et al., 2018], mis aitab andmetel luua patsiendipõhiseid spetsiifilisi ennustusmudeleid kasutades masinõppe algortimide laia assortiid.

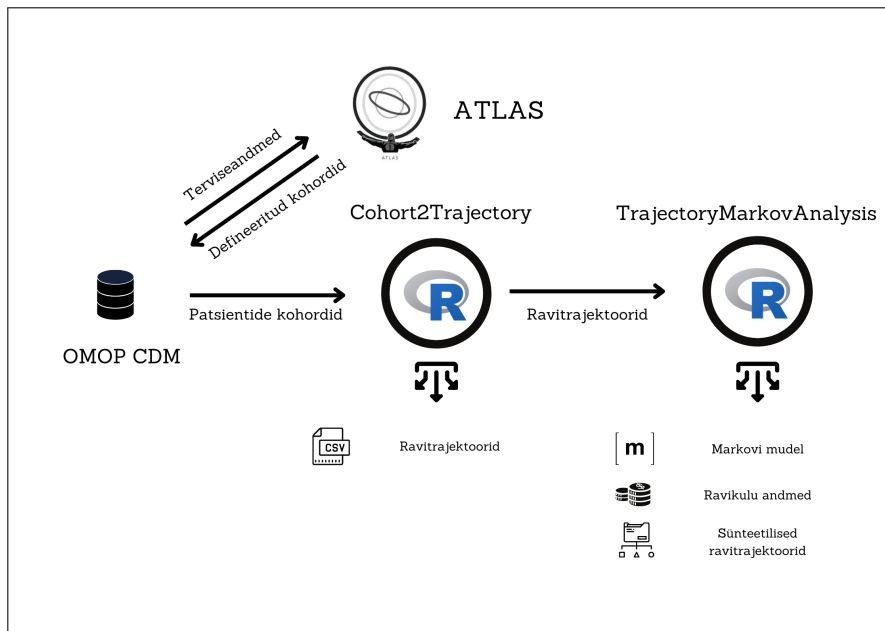
Lisaks OHDSI halduses olevale pakettide arendamise toetamisele on igal huvitundjal võimalik luua eraldiseisvaid OMOP ühtsele andmemudelile toetuvaid pakette. Kahjuks ei ole OHDSI veel koostanud ühtset juhendit ega regulatsioone, kuidas selliseid pakette luua. Sellest hoolimata on OHDSI välised arendajad ja teadlased loonud tänu suurele jälgijaskonnale ja HADES-pakettide võimekusele ning eeskujule taolisi R-pakette. Eelnevalt loodud pakettides on siiski välja kujunenud teatavad head praktikad. Oleme suutnud tuvastada järgnevad:

1. paketi *DatabaseConnector*'i ja *SqlRender*'i kasutamine erinevate SQL dialektide toetamiseks;
2. paketti kokkuvõtva funktsiooni koostamine;
3. kasutajale tulemuste kataloogi sätestamise valiku tekitamine;
4. juhul, kui kasutatakse kohorte, siis nende JSON ja SQL kuju salvestamise võimalus;
5. juhul, kui kasutajale peab määrama parameetrite väärtusi, siis need salvestatakse;
6. juhul, kui luuakse ühendus andmebaasiga, siis seda hõlmav kood on muust koodist eraldatud;
7. juhul, kui andmebaasi tekitatakse ajutisi tabeleid, siis need kustutatakse peale paketi kasutamist;
8. juhul, kui kasutatakse andmebaase, siis testitakse selle funktsioone Eunomia [Schuemie, 2020] andmebaasil.

Eelnevat arvesse võttes oleme antud magistritöö koostamisel välja arendanud kaks R-paketti, mis toetuvad OMOP ühtsele andmemudelile.

4 Implementatsioon

Käesolevas peatükis anname ülevaate magistritöö käigus loodud R-pakettidest. Magistritöö koostamisel on välja arendatud kaks R-paketti *Cohort2Trajectory* (<https://github.com/HealthInformaticsUT/Cohort2Trajectory>) ja *TrajectoryMarkovAnalysis* (<https://github.com/HealthInformaticsUT/TrajectoryMarkovAnalysis>). Pakettidest esimene koostab OMOP andmemudelit ja ATLAS-tarkvara kasutades patsientide ravitrajektorid. Teine valminud pakett võtab sisendiks esimese paketi väljastatud ravitrajektorid ning väljastab kasutajale trajektoridelt õpitud Markovi mudeli ning vajaliku info kulutõhususe analüüsi tegemiseks. Töö tulemusena valminud pakettide töövoog on toodud joonisel 3.



Joonis 3. Loodud pakettide rakendamise töövoog

4.1 R-pakett *Cohort2Trajectory*

Paketi *Cohort2Trajectory* eesmärk on tekitada patsiendi meditsiinilisele ajaloole toetudes defineeritud seisunditest ravitrajektor. Paketti saab kasutada OMOP ühtsele andmemudelitele ligipääsu olemasolul, lisaks on vaja andmete sisselugemiseks defineeritud seisundeid, mis on loodud ATLAS-tööriistas või eelnevalt salvestatud JSON-kuul. OMOP ühtne

andmemudel peab olema talletatud ühes järgnevas andmebaasihalduris: Microsoft SQL Server, Oracle, PostgreSQL, Amazon RedShift, Impala, IBM Netezza, Google BigQuery, Microsoft PDW, Apache Spark või SQLite. See nimistu tuleneb R-paketi *SqlRender* võimekusest renderdada parametrizeeritud SQL koodi ning tõlkida seda nendesse erinevatesse SQL dialektidesse [Schuemie ja Suchard, 2022b]. Kasutades defineeritud patsientide kohorte saab *Cohort2Trajectory* paketi tekitada vaatlusaluste isikute ravitrajektoore, mis kirjeldavad seisundites olemist (ajaintervall) ning säilitavad nende vaheliste üleminekute ajalise järjestuse. Arvestades paketi kasutajate sihtrühma, on pakett ingliskeelne.

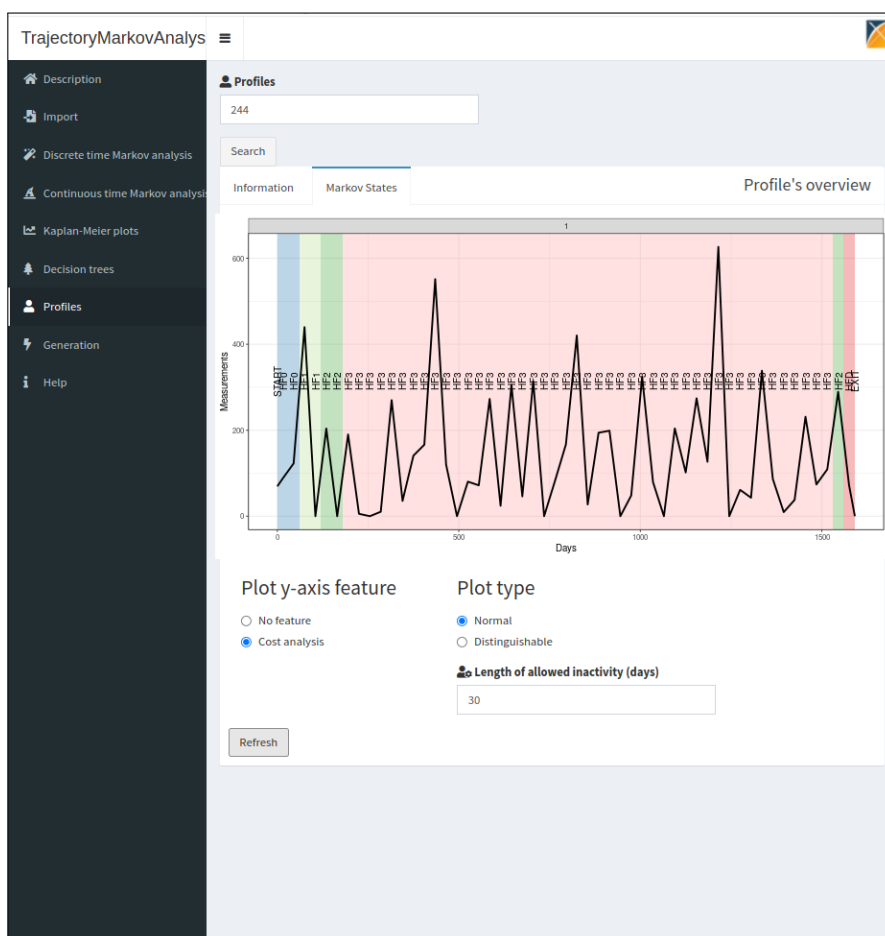
4.1.1 Paketi sisu

Paketi koostamiseks on kombineeritud mitmete R-pakettide funktsionaalsuseid. Kokku on kasutatud 22 varem koostatud paketi võimekust, nendest kuus on OHDSI hallatavad paketid. Pakett koosneb 27 erinevast funktsioonist, millest kasutajad saavad kasutada kahte. Need kaks funktsiooni kombineerivad endas kõikide ülejäänud funktsioonide rakendamise.

Paketti *Cohort2Trajectory* on võimalik kasutada kahel erineval viisil. Esiteks on võimalik pakett avada kasutajaliidesega (vt joonis 4). Kasutajaliides on arendatud välja kasutades R-paketi *shiny* versiooni 1.7.1 [Chang et al., 2021]. Kasutajaliides koosneb kaheksast vahelehest, milleks on:

1. kirjeldus (ingl *description*);
2. importimine ATLAS'st (ingl *import via ATLAS*);
3. importimine JSON'dest (ingl *import via JSONs*);
4. statistika (ingl *statistics*);
5. prioritseerimine (ingl *prioritization*);
6. trajektoolid (ingl *trajectories*);
7. profiilid (ingl *profiles*);
8. abi (ingl *help*).

Kirjelduse ja abi vahelehed on koostatud kasutaja juhendamiseks kasutajaliidese funktsioonide rakendamisel. Vaheleht "Importimine ATLAS'st" (vt lisa 1) laseb kasutajal märkida ära kõikvõimalikud ATLAS-tarkvaras defineeritud patsientide kohordid, mida



Joonis 4. Pakettide *Cohort2Trajectory* ja *TrajectoryMarkovAnalysis* kasutajaliidese vorm

trajektoore koostamisel kasutatakse. Eraldi tuleb märkida sihtkohort ja seisundite kohordid. Samuti saab kasutaja kohortide nimetusi muuta. Vahelehel "Importimine JSON'dest" (vt lisa 2) saab kasutaja defineerida kohordid JSON kujul. Juhul, kui kasutaja on juba varasemalt JSON kujul kohordid defineerinud, saab ta need kasutajaliidese üles laadida. "Statistika" (vt lisa 3) leheküljel kuvatakse kasutajale imporditud kohortide ja koostatud trajektoore kohta statistika. Leheküljel "Prioritiseerimine" (vt lisa 4) saab kasutaja mugavalt "drag and drop" viisil tekitada sisseloetud kohortidest prioritiseeritud järjekorra. Leheküljel "Trajektoolid" saab kasutaja defineerida erinevaid parameetreid trajektoore defineerimiseks (vt lisa 5). Leheküljel "Profiilid" saab kasutaja vaadata eraldiseisvalt isikute trajektoore (vt lisa 13).

Teiseks on paketti võimalik kasutada käsurealt. See variant jaguneb omakorda kaheks: uuringu defineerimine ja salvestatud uuringu taaskordne teostamine. Võimalus kasutada paketti käsurealt tekitab variandi liita selle rakendamine teiste OHDSI põhiste R-pakettide töövoogudega. Selle tõttu võib *Cohort2Trajectory* paketti potentsiaalselt nimetada ka OHDSI uuringute töövoogu toetavaks paketiks. Paketi väljundit saab kasutada näiteks Markovi mudelite, juhumetsade või neurovõrkude koostamiseks.

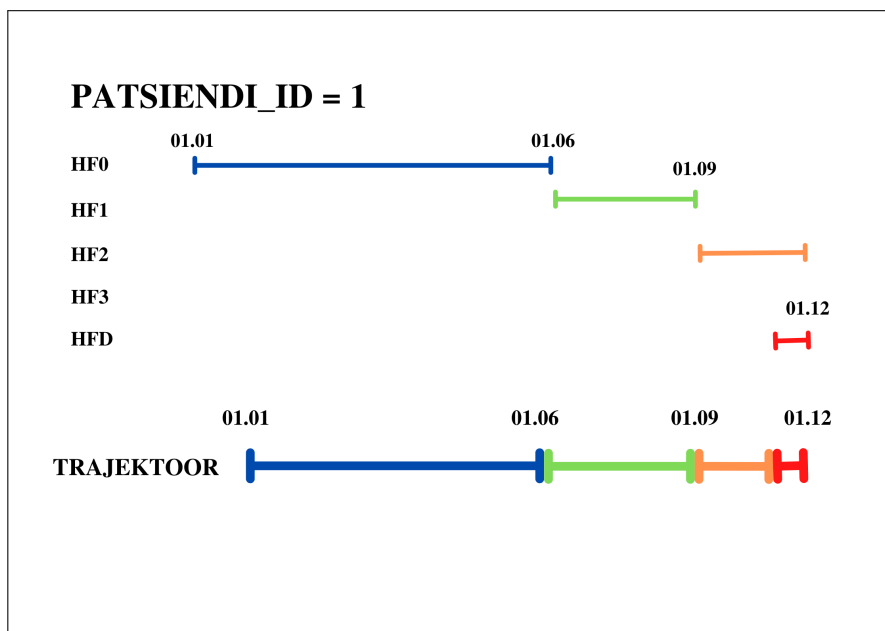
4.1.2 Trajektooride valmistamine

Järgnevalt kirjeldame paketi *Cohort2Trajectory* abil trajektooride loomist. Trajektooride moodustamisel on esimeseks ülesandeks defineerida sihtkohort (ingl *target cohort*), mis koosneb kõikidest kaasatavatest patsientidest ja nende vaatlusperioodidest (igal patsiendil erinev). Juhul, kui sihtkohort ei ole veel defineeritud JSON failina, siis saab seda teha kasutades ATLAS-rakendust (vt lisa 6). Sarnaselt sihtkohordile tuleb defineerida ka seisundite kohordid. Olenevalt kasutusviisist tuleb defineeritud kohordid importida ning seda siis kas andmebaasist või JSON'i kataloogist. Kasutajaliidese kasutamisel on imporditud andmete esialgset statistikat võimalik näha vastaval vahelehel.

Peale andmete importimist järgneb andmete puhastamise faas. Andmete puhastamisel eemaldatakse patsiendid, kes sihtkohorti ei kuulu ja ei rahulda kasutaja defineeritud kitsendusi. Kasutajal on võimalus defineerida kolm parameetrit, milleks on obligatoorsed seisundid, neelavad seisundid ja sihtkohordi vaatlusperioodist välja jäävate seisundite kaasamine. Obligatoorsete seisundite kaasamine tähendab, et kõik subjektid, kelle andmete hulgas pole antud seisundeid, eemaldatakse. Neelavate seisundite kaasamisel katkestatakse trajektooriid nendesse seisunditesse jõudmisel. Sihtkohordi vaatlusperioodist välja jäävate seisundite kaasamisel ei sõeluta vastavaid seisundeid välja. Juhul, kui need on keelatud, siis jäetakse pärast sihtkohordi vaatlusperioodi lõppemist esinevad seisundid trajektooriga välja. Viimane ei kehti juhul, kui toimub sihtkohordi ja seisundi kohordi vaatlusperioodide kattumine. Sel juhul jääb seisundi kohorti andmetes esindama perioodide ühisosa.

Trajektoore, mis tekivad, on võimalik koostada erinevatel viisidel. Esiteks saab valida, sõltuvalt tulevase analüüsi eesmärkidest, seisundite kestuse parameetreid. Kasutajal on võimalik tekitada diskreetseid trajektoore, kus kõikide seisundite pikkus on defineeritud ühe kindla arvu päevadega, ja pidevaid trajektoore, kus seisundite kestus vastab täpselt andmebaasist päritud andmetele (ehk iga eraldiseisev seisundis olek võib kesta t -päeva, $t \in \mathbb{R}_+$). Teiseks tuleb valida seisundite kattuvuste esinemisel meetod, millega konflikt lahendada. *Cohort2Trajectory* pakub selleks kolme erinevat võimalust, mida on kirjeldatud allpool.

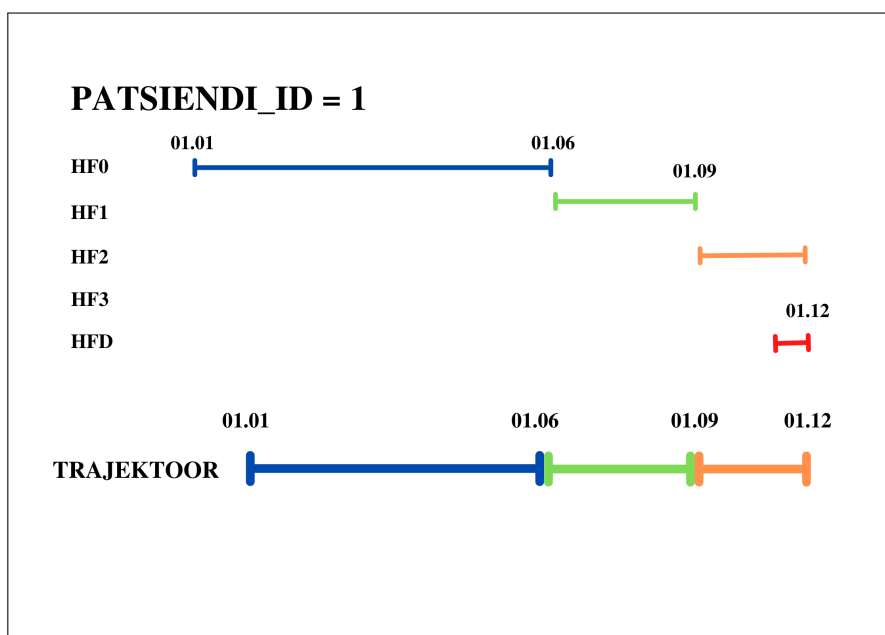
Esiteks on võimalus defineerida seisundite prioriteedid. Olukorras, kus kaks või enam seisundit esinevad samas ajaintervallis, jääb trajektoori alles seisund, mis on kõrgema prioriteediga. Nii diskreetse kui pideva trajektoori puhul esindab vaatlusalust ajaintervalli kõrgeima prioriteediga esinenud seisund (vt joonis 5).



Joonis 5. Prioriteedipõhine järjestus (järjestus on defineeritud kõrgemalt madalamale HFD, HF3, HF2, HF1 ja HF0)

Teiseks on võimalus kattuvuskonflikti lahendada esinemispõhise järjekorraga. Olukorras, kus kaks või enam seisundit esinevad samas ajaintervallis, jääb trajektoori alles seisund, mis ilmub varem. Diskreetse ravitrajektoori puhul tähendab see seda, et kui n -päevases vaatlusperioodis on esindatud mitu erinevat seisundit, siis trajektoori ajaintervalli jääb esindama vaatlusperioodis esimesena esinenud seisund. Pideva ravitrajektoori korral jääb mitme seisundi samaaegselt esinemise puhul alles täispikkuses esimesena esinenud seisund (vt joonis 6). Olukorras, kus mitu seisundit algavad samal ajahetkel, jääb alles kõrgema prioriteediga seisund.

Kolmandaks on võimalus kattuvuskonflikti lahendada suurema kattuvusega seisundi eelistamisega. Olukorras, kus kaks või enam seisundit esinevad samas ajaintervallis, jääb trajektoori alles seisund, mis kattub vaatlusaluse ajaintervalliga enim. Siin on võimalik vaid diskreetse trajektoori juht (vt joonis 7). Olukorras, kus mitu seisundit on esindatud kõrgeima arvu kattuvate päevadega, jääb alles kõrgema prioriteediga seisund.



Joonis 6. Esinemispõhine järjestus

Trajektooride koostamisel võib tekkida olukord, kus patsiendi seisundites olek katkeb mingiks ajaks. See tähendab, et mingis ajaintervallis ei kuulu patsient mitte ühtegi seisundi kohorti. Me lahendame selle olukorra tähistades vastava intervalli seisundiga "kohordist väljas" (ingl "out of cohort"), tagades sellega trajektooride pidevuse. Lisaks on igale valminud trajektoorile lisatud kaks seisundit tähistamaks trajektoori algust ja lõppu (vastavalt "start" ja "end"). Alguspunkti tähistav seisund on dateeritud sihtkohorti sisenemise päevaga. Lõpp-punkti tähistav seisund on dateeritud patsiendi viimase seisundi lõppkuupäevale järgneva päevaga ning on loodud juhuks, kui uuringust puuduvad neelduvad seisundid. Need kaks seisundit on hiljem vajalikud R-pakettide funktsioonide kasutamiseks.

Kuna paketi funktsioonid eeldavad mitmete *for*-tsüklite kasutamist, siis on ressur-sinõudlikumad funktsioonid kirjutatud R asemel C++ programmeerimiskeelt kasutades. Programmeerimiskeeli R ja C++ saab hõlpsasti kombineerida kasutades paketti *Rcpp* [Eddelbuettel ja Balamuta, 2018].

Pärast trajektooride koostamist väljastatakse komakohtadega eraldatud fail (vt lisa 7). Kõnealusel failis on kaheksa tulp, mis kirjeldavad ridade kaupa patsiendi kuuluvust seisunditesse. Tulpades olev info:

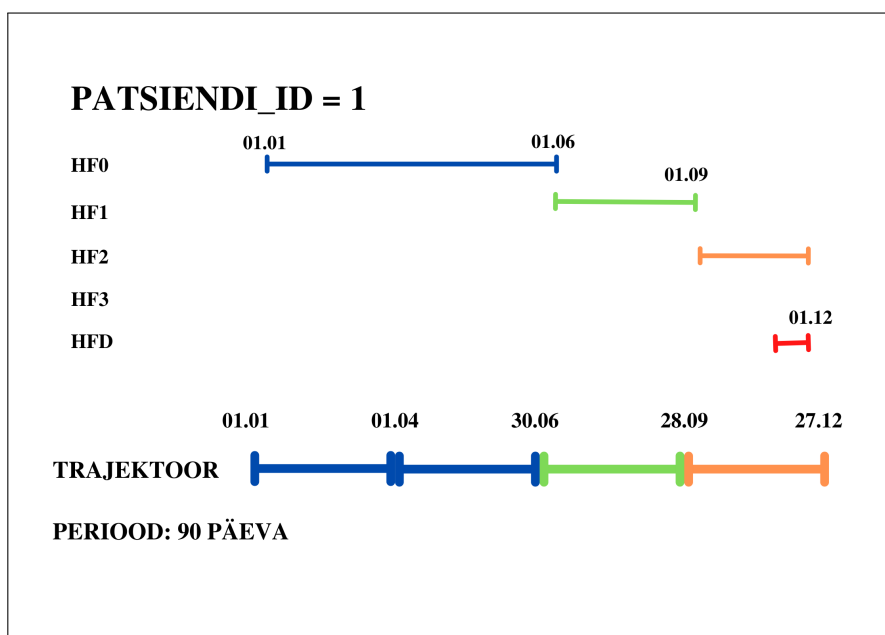
1. patsiendi identifikaator;
2. seisundi nimetus;
3. seisundisse sisenemise kuupäev;
4. seisundist väljumise kuupäev;
5. möödunud aeg trajektoori algushetkest (aastates);
6. patsiendi sugu;
7. patsiendi vanus antud ajahetkel;
8. seisundi identifikaator;

4.1.3 Võimalikud arendused

Antud kujul katab R-pakett *Cohort2Trajectory* trajektooride tekitamise põhilised vajadused. Kindlasti on paketil rohkem potentsiaali kui praegune funktsionaalsus suudab katta ning seega on põhjust paketti edasi arendada. Paketi testijad on toonud välja järgmisi võimalike täiendusi:

- võimalus tekitada ühele patsiendile mitu trajektoori vastavalt mõne tingimuse täitumisele;
- võimalus eristada seisundeid väiksema ajaintervalliga (hetkel on kõige lühem periood üks päev);
- võimalus defineerida lubatud seisunditevahelised üleminekud;
- luua rohkem võimalusi kattuvuskonfliktide lahendamiseks.

Antud pakett standardiseerib andmestiku koostamise ja OMOP andmebaasiga suhtlemise sammud ravitrajektooride uurimisel. Lisaks Markovi ahelate tööriistale toetuvad sellele ka üks trajektooride visualiseerimise tööriist [Pajusalu, 2022] ja ravitrajektooride sarnasuse algoritme uuriv bakalaureusetöö. Et julgustada teisi arendajaid seda paketti kasutama, plaanime kandideerida OHDSI HADES’ e pakettide gruppi.



Joonis 7. Kattuvuspõhine järjestus

4.2 R-pakett *TrajectoryMarkovAnalysis*

Paketi *TrajectoryMarkovAnalysis* eesmärk on modelleerida patsientide ravitrajekte Markovi mudelitega. Lisaks üleminekutõenäosustele hindab pakett seisundipõhiseid kulusi. Paketi kasutajaliideses on Markovi mudelitele võimalik koostada Kaplan-Meieri graafikuid ning otsustuspuud. Töö eesmärkidest lähtuvalt saab seda paketti kasutada OMOP ühtsele andmemudelile ligipääsu olemasolul, lisaks on vaja patsientide ravitrajekte paketi *Cohort2Trajectory* väljastatud kujul. OMOP ühtne andmemudel peab olema talletatud mõnes R-paketi *SqlRender* toetatud SQL dialektis. Sarnaselt paketile *Cohort2Trajectory* on see pakett ingliskeelne.

4.2.1 Paketi sisu

Paketi koostamiseks on kombineeritud mitmete R-pakettide funktsionaalsuseid. Kokku on kasutatud 27 varem koostatud paketi võimekust, nendest kolm on OHDSI hallatavad paketid. Pakett koosneb 31 erinevast funktsioonist, millest kasutajad saavad kasutada nelja. Need neli funktsiooni kombineerivad endas kõikide ülejäänud funktsioonide kasutamist.

Paketti *TrajectoryMarkovAnalysis* on võimalik kasutada kahel erineval viisil – kasutajaliidesega (vt joonis 4) või käsureaal funktsioonide väljakutsumisega. Kasutajaliides on arendatud välja kasutades R-paketti *shiny* versiooni 1.7.1 [Chang et al., 2021]. Kasutajaliides koosneb üheksast vahelehest, milleks on:

1. kirjeldus (ingl *description*);
2. importimine (ingl *import*);
3. diskreetse ajaga Markovi ahelad (ingl *discrete time Markov analysis*);
4. pideva ajaga Markovi ahelad (ingl *continuous time Markov analysis*);
5. Kaplan-Meieri graafikud (ingl *Kaplan-Meier plots*);
6. otsustuspuud (ingl *decision trees*);
7. profiilid (ingl *profiles*);
8. genereerimine (ingl *generation*);
9. abi (ingl *help*).

Vahelehtedel "Abi" ja "Kirjeldus" antakse kasutajale infot paketi kasutamise kohta. Kasutaja saab trajektooride andmed üles laadida leheküljel "Importimine", mille järel kuvatakse andmetes leiduvad trajektoorid (vt lisa 8).

Leheküljel "Diskreetse ajaga Markovi ahelad" (vt lisa 9) saab tekitada imporditud andmetele tuginedes Markovi mudeli. Kasutaja saab valida, millistele OMOP andmemudeli *Cost* tabeli domeenide (ravimid, visiidid, protseduurid, seadmed, mõõtmised, uuringud, proovid) maksumustele rahalist analüüsi tehakse. Samuti on kasutajal võimalik mõningaid seisundeid uuringust välja jätta. Kasutajale kuvatakse arvutatud üleminekumaatriks, trajektooride algseisundid ning nende kuluanalüüs, seisunditepõhine kuluanalüüs ning samuti näidatakse seisundite kulude jaotust. Kasutaja saab kohandada üleminekumaatriksit sobiva arvu tsüklite jaoks.

Pideva ajaga Markovi ahelad koostatakse leheküljel "Pideva ajaga Markovi ahelad" (vt lisa 10) kasutades patsientide trajektoore ja R-paketti *msm* [Jackson, 2011]. Paketi *msm* funktsioonid toetavad lisaks pideva ajaga Markovi mudelite loomisele erinevate kovariaatide kasutamist. Lisaks eelnevale on *msm* paketti implementeeritud paralleelarvutused, seega on see võrdlemisi kiire. Kasutaja saab valida, millistele domeenide maksumustele rahalist analüüsi tehakse. Samuti on kasutajal võimalik mõningaid seisundeid uuringust

välja jätta. Kasutaja ülesandeks on defineerida pideva Markovi mudeli intensiivsusmaatriksi Q võimalikult lähedane kuju. Muutes maatriksi Q i -nda rea ja j -nda veeru elemendi võrdseks nulliga, annab see pakatile teada, et seisundist i seisundisse j on üleminek keelatud. Kasutajale kuvatakse arvutatud üleminekumaatriks, trajektooride algseisundid ning nende kuluanalüüs, seisunditepõhine kuluanalüüs ning samuti näidatakse seisundite kulude jaotust. Kasutaja saab kohandada üleminekumaatriksit t ($t \in \mathbb{R}_+$) arvu päevade jaoks.

Lisaks Markovi ahelatele kasutatakse patsientide ravitrajektooride uurimiseks otsustuspuud [Kobelt, 2013]. Selleks lisasime kasutajaliidesesse vahelehe "Otsustuspuud" (vt lisa 12), kus kasutaja saab otsustuspuu kihtide kaupa ülesse ehitada ning näha seisunditevaheliste liikumiste tõenäosuseid.

Selleks, et visualiseerida erinevate seisundite üleminekuid sihtkohordis, lisasime võimaluse luua kohordipõhiseid Kaplan-Meieri graafikuid (elukestuskõveraid). Nende kasutamine lubab kasutajal saada infot patsientide elulemusfunktsiooni kohta. Elulemusfunktsioon annab infot seisundite vahelise liikumise intensiivsuse ja ajakulu vahel. Samuti saab võrrelda, kas kaks erinevat elulemusfunktsiooni on statistiliselt erinevad. Leheküljel "Kaplan-Meieri graafikud" (vt lisa 11) saab kasutaja koostada Kaplan-Meieri graafikuid kõigile võimalikele andmetes leiduvatele üleminekutele. Elukestuskõverate implementeerimisel toetusime suuresti R-paketile *survminer* [Kassambara et al., 2021]. Lisaks on implementeeritud võimalus võrrelda erinevaid vanuserühmi, mille statistilist erinevust võrreldakse astaktestiga.

Kasutajaliidesesse on implementeeritud erinevate trajektooride võrdlemiseks logaritmilise astaktesti (ingl *log-rank test*) rakendamine. Seda saab kasutada nii vaadeldud andmete omavaheliseks (erinevate vanuserühmade vahel) võrdlemiseks kui Markovi mudelite põhjal genereeritud trajektooride valideerimiseks (võrdluses reaalsete patsientide trajektooridega). Järgnevalt anname astaktesti statistikust ülevaate põhinedes Dave Schoenfeldi käsitlusele [Schoenfeld, 1981].

Üldsust kitsendamata laiendame seda elukestuskõverate statistikut ka antud töös kirjeldatavatele seisunditevaheliste üleminekute kõveratele. Oletame, et meil on n patsienti, kes on jaotatud kahe erineva grupi 0 ja 1 vahel. Olgu meie alternatiivseks hüpoteesiks, et gruppide 0 ja 1 vahel on erinevad kõverad seisundite a ja b vahel. Olgu $j = 1, \dots, n$ patsiendi number ja koosnegu hulk D patsientidest, kes liikusid seisundist a seisundisse b . Lisaks tähistagu t_j ajahetke, mil patsient j liikus seisundist a seisundisse b . Olgu, et $X_j = 0, 1$ defineerib patsiendi j gruppi kuuluvust. Olgu ajahetk t ja tähistagu arv $n_i(t)$ patsiente, kes kuuluvad gruppi i ja pole veel liikunud seisundist a seisundisse b . Lisaks

defineerime veel, et $p(t) = n_1(t)/(n_0(t) + n_1(t))$. Astaktesti statistikuks on

$$S = \sum_{j \in D} \frac{(X_j - p(t_j))}{\sum_{j \in D} p(t_j(1 - p(t_j)))^{\frac{1}{2}}}.$$

Selle statistiku rakendamine genereeritud andmete valideerimiseks on õigustatud. Lisaks genereeritud andmetel tekkiva Markovi mudeli väärtustele hindab see mudeli võimekust kirjeldada üleminekuks kuluvat aega (intensiivsust). Sarnast lähenemist on kasutatud emakakaelavähki kirjeldava Markovi ahela valideerimiseks [Soper et al., 2020].

Tänapäeval on tervishoiusüsteemides tõusnud nõudlus kaasaegsete tehisintellekti tehnoloogiate kasutuselevõtu järele. Paljud neist tehnoloogiatest kasutavad patsientide ajaloolisi terviseandmeid, et luua võimsaid ennustumudeleid, mida saab kasutada haiguste mõistmise ja diagnoosimise parandamiseks. Selliste ennustumudelite loomisel on patsientide andmete privaatsus üheks läbivaks probleemiks. Selle tõttu on sünteetilistel tervisandmetel andmeteaduses tähtis roll. Meie loodud tööriist omab funktsiooni, mis genereerib mudeli põhjal sünteetilisi patsientide ravitrajektoore. Kasutajaliideses on lehel "Genereerimine" (vt lisa 14) võimalik tekitada sünteetilisi trajektoore koostatud Markovi mudelite ja Monte Carlo meetodite põhjal.

Erinevalt paketist *Cohort2Trajectory* tekitab paketi *TrajectoryMarkovAnalysis* töövoog andmebaasi ajutisi tabeleid. Need kõik luuakse patsientide seisundite uurimiseks ning nende maksumuste analüüsi kiirendamiseks. Kõik tabelid koostatakse kasutaja määratud andmebaasi skeemi ning kustutatakse soovi korral pärast igakordset paketi kasutamist.

4.2.2 Võimalikud arendused

R-pakett *TrajectoryMarkovAnalysis* suudab lahendada mitmeid ravitrajektooriga seotud ülesandeid, kuid kõiki neid üsna pinnapealselt. Väga spetsiifiliste funktsioonide rakendamine võib aga omakorda muuta paketi kasutamise segasemaks ja vähem universaalseks. Siiski pakume välja mõningad võimalikud arendused.

- Paketi *msm* funktsioonid toetavad lisaks pideva ajaga Markovi mudelite loomisele erinevate kovariaatide kasutamist. Seega oleks võimalik võrrelda erinevaid patsientide gruppe.
- Kuigi Markovi protsessid on mäluta protsessid, siis Markovi mudeleid saab tekitada selliseski vormis, kus ühe seisundi asemel vaadeldakse kahte (või enam) eelmist seisundit. Kuna võimekus selliseid mudeleid koostada on olemas, sobiks nad ka *TrajectoryMarkovAnalysis* paketi funktsioonide hulka.

- Kuna OHDSI kogukonna üheks eesmärgiks on uuringute jagamine ja tulemuste võrdlemine, siis oleks tarvis arendada tulemuste võrdlemiseks (andmebaasivaheliseks) kasutajaliides.
- Hetkel on võimalik saada andmeid maksumuste kohta sõeludes välja domeene *Cost* tabelist ning agregeerides need mingis teatud ajaperioodis. Vaadates tulevikku tundub oluline siduda maksumused kindlate oluliste meditsiiniliste toimingutega.
- Kasutajaliides toetab nii Kaplan-Meieri graafikute kui otsustuspuude loomist. Nende vahelehtede kasutajasõbralikkust tuleks veel arendada.

5 Näidisuuring: südamepuudulikkusega patsientide ravitrajektorid

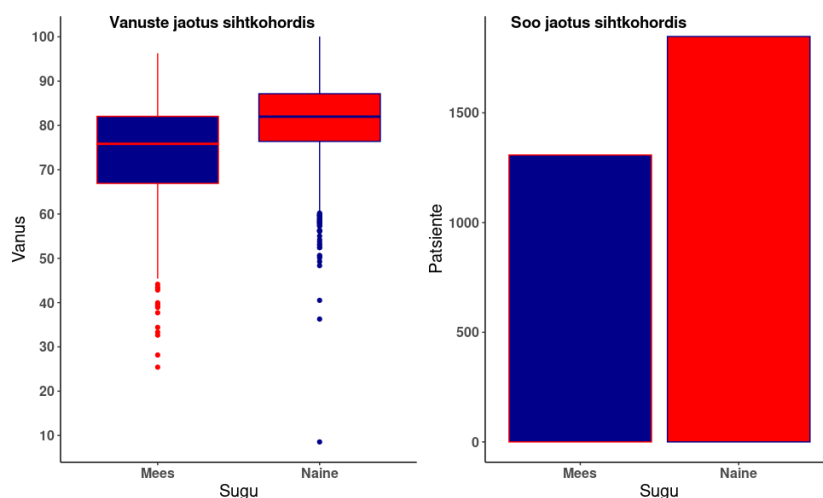
Käesolevas peatükis näitame eelmainitud pakettide funktsionaalsuseid. Selle läbiviimiseks toetume ja proovime reprodutseerida Praveen Thokala [Thokala et al., 2020] artiklis esitatud uuringut, kus uuriti Suurbritannia südamepuudulikkusega patsiente. Südamepuudulikkusega patsientide ravimine moodustab enamikes arenenud riikides 1%–2% aastasest tervishoiu eelarvest, seda iseloomustavad kõrge sagedus ning suremus. On leitud, et telemonitooringu kaasamine patsientide ravil aitab tõsist haigestumist ennetada ja selle tulemusena stabiliseerida patsientide seisundit ning ära hoida hospitaliseerimisi [Thokala et al., 2020]. Thokala et al. uurisid, kas telemonitooringu kasutamine võrreldes tavalise raviga osutub piisavalt efektiivseks, et tervishoiusüsteem võiks selle lähenemise kasutusele võtta. Sellise uuringu läbiviimiseks kasutasid nad diskreetse aja Markovi mudeleid, mille parameetrid arvutati empiirilistelt ravitrajektoridelt.

Antud uuringu läbiviimiseks jagati patsientide jälgimisperiood 30-päevasteks intervallideks, mis kirjeldasid patsiendi terviseseisundit. Seisundid eraldati eelneva aasta jooksul südamepuudulikkuse tõttu haiglaviisiidil käidud kordade arvu põhjal. Seega esimene seisund (edaspidi HF0) kirjeldas patsienti, kes ei olnud eelneva aasta jooksul kordagi südamepuudulikkuse tõttu haiglas käinud. Teine seisund (edaspidi HF1) kirjeldas patsienti, kes oli ühe korra eelneva aasta jooksul südamepuudulikkuse tõttu haiglas käinud. Kolmas seisund (edaspidi HF2) kirjeldas patsienti, kes oli kaks korda eelneva aasta jooksul südamepuudulikkuse tõttu haiglas käinud. Neljas seisund (edaspidi HF3) kirjeldas patsienti, kes oli kolm või rohkem korda eelneva aasta jooksul südamepuudulikkuse tõttu haiglas käinud. Viies seisund (edaspidi HFD) kirjeldas patsiendi surma. Sihtkohordis vaadeldi inimesi, kellel oli esinenud vähemalt üks südamepuudulikkusega seotud hospitaliseerimine, kuni nende surmani.

5.1 Andmed

Näidisuuringu läbiviimiseks on kasutatud Eesti Haigekassa andmeid, mis on viidud OMOP ühtse andmemudeli formaati. Andmeid oli kasutada kokku 386557 patsiendi kohta, nendest kuulus sihtkohorti 3224 inimest (võrreldavas uuringus 2825 inimest). Patsientidest 41,4% olid mehed ning 58,6% naised (vt joonis 8). Keskmised vanused olid vastavalt 73,93 ja 80,93 aastat. Kõikide patsientide jälgimisperioodiks kujunes ajavahemik 2017. aasta jaanuarist 2022. aasta jaanuarini.

Kasutades paketti *Cohort2Trajectory* konverteerisime info kohortidesse kuulumisest



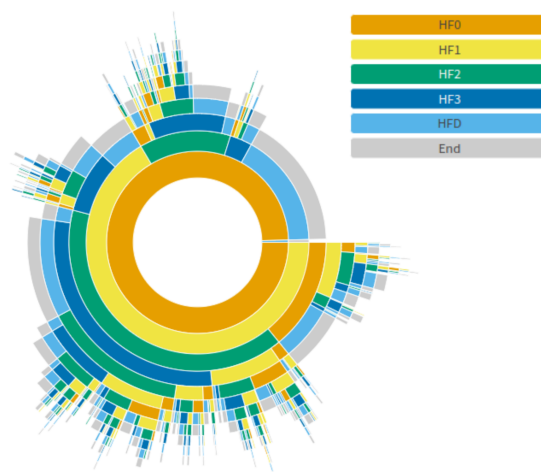
Joonis 8. Patsientide soo ja vanuse jaotused sihtkohordis

ravitrajektoorideks. Defineerisime ATLAS-tarkvaras kuus kohorti: ühe sihtkohordi ja viis seisundi kohorti. Nende kohta võib näha ülevaadet lisas 3, kus on toodud mõningane statistika ja kattuvuskaart. Lisas 3 esitatud kattuvuskaardilt võib näha, et defineeritud seisundid kattuvad mõningal määral. Enim kattuvad teiste seisunditega seisundid HFD ja HF0. See on oodatav tulemus, sest mõlemad seisundid on sihtkohorti kuulumiseks vajalikud.

Selleks, et reprodutseerida näidiseks toodud uuringut defineerisime trajektooride moodustamise järgnevalt:

- diskreetse aja trajektooride tüüp;
- seisundi pikkus 30 päeva;
- prioriteedipõhine valik;
- neelav seisund HFD ;
- obligatoorne seisund HFD;
- lubatud seisundid väljaspool sihtkohordi vaatlusperioodi.

Kasutades paketti *Cohort2Trajectory* koostasime vaatlusaluste patsientide ravitrajektooreid. Patsientide liikumised erinevate seisundite vahel vaatlusperioodil on toodud joonisel 9. Jooniselt võib näha, et pea kõik subjektid alustavad seisundist HF0 ning kõik ravitrajektooreid lõppevad surmaga.



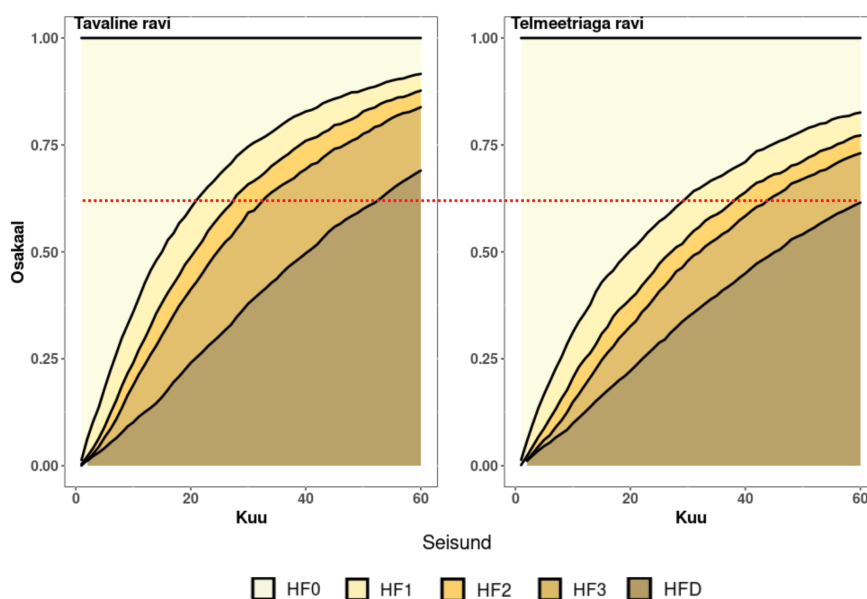
Joonis 9. Seisundite kronoloogiline järgnevus

5.2 Tulemused

Kasutades paketi *Cohort2Trajectory* koostatud diskreetseid trajektoore tekitasime paketi *TrajectoryMarkovAnalysis* abil diskreetse aja Markovi mudeli ning võrdleme tulemusi Suurbritannias saadud tulemustega. Samuti uurisime, kas Eestis oleks telemeetria rakendamine otstarbekas.

Ravitrajektoorida analüüsi tulemuseks oli Markovi mudel (vt tabel 4), mis kirjeldab südamepuudulikkusega patsientide seisundite HF0, HF1, HF2, HF3 ja HFD vaheliseid üleminekutõenäosuseid. Markovi mudelist on näha, et enamasti patsiendid ühe kuu jooksul seisundit ei vaheta. Seisundid HF0, HF1, HF2 ja HF3 on ebaolulised seisundid ning HFD on neelav seisund. Kõige suurem võimalus (3,7%) järgmisel kuul surra on olles seisundis HF3.

Eesti ja Suurbritannia andmetel koostatud Markovi mudelite (vt tabel 4) vahel on palju sarnasusi, kuid esineb ka erinevusi. Näiteks seisundi HF3 ja HFD vaheline liikumine on Suurbritannia kohordis ligi 3 korda tõenäolisem. Samas Eesti patsientidel progresseerub haigus märgatavalt kiiremini, seisundist HF0 seisundisse HF3 liikumise tõenäosus läbides järjest seisundeid HF0, HF1, HF2 ja HF3 on ligi 280 korda tõenäolisem kui Suurbritannia patsientidel. Siin võivad rolli mängida mitmed asjaolud. Esiteks ei ole teada Suurbritannia uuringusse kaasatud patsientide vanus-sooline jaotus. Selle olemasolul oleksime saanud oma sihtkohordile rohkem tingimusi panna. Teiseks võib õpitud mudeli parameetreid mõjutada nii patsientide aktiivsus arsti juures käimise suhtes kui südamepuudulikkuse diagnoosi märkimine hospitaliseerimise põhjusena. Kolmandaks tuleb arvestada sellega,



Joonis 10. Seisundite jaotus jälgimisperiodil mudelitest genereeritud andmetel

et mõlemasse uuringusse kaasatud subjektide arv on üsna madal.

Lisaks üleminekute tõenäosustele pärisime pakettist *TrajectoryMarkovAnalysis* erinevate kulutustega seotud suuruseid (vt lisa 9). Kõikide patsientide keskmine trajektoori läbimise kulu oli 9505€. Mediaankulu oli 7375€. Keskmiselt kulus kuu aja jooksul Haigekassal patsientide peale vastavalt 243,3€ (seisundis HF0), 266,1€ (seisundis HF1), 275,1€ (seisundis HF2) ja 315,3€ (seisundis HF3).

Selleks, et võrrelda südamepuudulikkusega patsientidele hetkel rakendatavat ravi variandiga, kus kaasatakse ravisse telemeetria, pidime leidma alternatiivse ravi Markovi mudeli. Selleks kasutame me riskitiheduse suhet (ingl *hazard ratio*). Riskitiheduse suhet (HR) saab arvutada valemiga

$$HR = \frac{\text{uue ravi risk}}{\text{vana ravi risk}},$$

kus riskidena võib vaadelda üleminekutõenäosusi [Spruance et al., 2004]. Varasematest metaanalüüsides on välja tulnud, et telemeetria rakendamisel (inimeste tööajal) on riskitiheduse suhe igasugusele suuremisele 0,75 ja hospitaliseerimisele 0,76 [Pandor et al., 2013].

Tabel 4. Südamepuudulikkusega patsientide Markovi ahelad tsükli pikkusega 1 kuu

Markovi mudel Eesti andmetel					
Tervise seisund	HF0	HF1	HF2	HF3	HFD
HF0	0,9457	0,0371	0,0075	0,0019	0,0146
HF1	0,0560	0,7807	0,1167	0,0228	0,0238
HF2	0,0150	0,0724	0,7027	0,1805	0,0294
HF3	0,0026	0,0073	0,0331	0,9197	0,0373
HFD	0,0000	0,0000	0,0000	0,0000	1,0000
Markovi mudel Suurbritannia andmetel					
Tervise seisund	HF0	HF1	HF2	HF3	HFD
HF0	0,981	0,004	0,000	0,000	0,015
HF1	0,068	0,872	0,017	0,000	0,043
HF2	0,004	0,094	0,777	0,041	0,084
HF3	0,000	0,006	0,095	0,777	0,122
HFD	0,000	0,000	0,000	0,000	1,000

Kasutades neid andmeid arvutasime alternatiivse ravi maatriksi, kus on rakendatud riskitiheduse suhteid vastavalt alltoodule.

1. Kõik seisundisse HFD viivad tõenäosused on korrutatud läbi suuremise riskitiheduse suhtega.
2. Kõik liikumised, mis tõstsid aastast hospitaliseerimiste arvu (HF1'st HF2'te, HF1'st HF3'me ja HF2'st HF3'e), on korrutatud läbi hospitaliseerimise riskitõhususe suhtega.
3. Uue maatriksi kõik read jagati läbi reasummadega, et kehtiks $\sum_j p_{ij} = 1$.

Telemeetria rakendamisel kehtiv Markovi mudel on toodud tabelis 5.

Kahe erineva ravi Markovi mudelite võrdlemisel jääb selgelt silma, et telemeetria kasutamisel progresseerub haigus aeglasemalt. Seega on paslik teha telemeetria rakendamisele kulutõhususanalüüs. Eesti Haigekassa on sätestanud hetkel kehtivaks ööpäevaringse telemeetria rakendamise hinnaks 64,03€ iga patsiendi kohta [Vabariigi Valitsus, 2022]. Selleks, et seda suurus kasutada, oletame, et tööajal telemonitooringu kasutamine maksab päevas $\frac{64,03 \cdot 9}{24} \approx 24,01€$. See tähendab igas kuus (arvestades 20 tööpäevaga) 480,22€ suurus täiendkulu. Arvutame järgnevalt, kui palju maksaks keskmiselt patsiendi ravi viie aasta jooksul Haigekassale vastavalt tema algseisundile (HF0, HF1, HF2 ja HF3). Tabelist 6 on näha, et mida rohkem haigus progresseerunud on, seda vähem see Haigekassale

Tabel 5. Südamepuudulikkusega patsientide Markovi ahel tsükli pikkusega 1 kuu telemeetria rakendamisel

Tervise seisund	HF0	HF1	HF2	HF3	HFD
HF0	0,9533	0,0283	0,0057	0,0014	0,0113
HF1	0,0584	0,8137	0,0912	0,0178	0,0189
HF2	0,0158	0,0764	0,7414	0,1428	0,0236
HF3	0,0026	0,0074	0,0334	0,9282	0,0284
HFD	0,0000	0,0000	0,0000	0,0000	1,0000

maksab. Seda põhjustab inimeste jõudmine surmaseisundisse HFD, mille kulu on 0 €. Lisas 9 toodud info kohaselt algasid meie andmetel 99,59% juhtudel ravitrajektorid seisundist HF0 ülejäänud seisundist HFD (patsiendid, kes surid esimese kuu jooksul). Tabelis 6 on toodud patsientide keskmised kulud Haigekassale sõltuvalt nende algseisundist ja ravimeetodist. Suuruste arvutamiseks on summeeritud iga kuu seisundite jaotus (kasutatud valemit 1) ning korrutatud läbi andmetest leitud seisundite maksumusega. Täiendkulu tõhususe määra arvutamiseks kasutame infot seisundi HF0 kohta tabelist 6 ning kasutame Thokala et al. uuringust leitud QALY'de vahe hinnangut 0,075. Eelnevast saame, et

$$TMM = \frac{4555,19}{0,075} \approx 60735,87\text{€/QALY},$$

mille tulemus on sarnane Suurbritannias tehtud uuringu täiendkulu tõhususe määrale (vastavalt 56316,94€/QALY). Kuivõrd Eestis loetakse ühe QALY piirmääraks 40000€ [Pokrovski et al., 2019], siis meie andmetest selgus, et telemeetria kasutamine südamepuudulikkusega patsientide ravil osutub selgelt liiga kalliks.

Tabel 6. Südamepuudulikkusega patsientide keskmine kulu Haigekassale sõltuvalt algseisundist

Algseisund	Kasutatav ravi (€)	Alternatiivne ravi (€)	Vahe (€)
HF0	1944,25	6499,44	4555,19
HF1	1829,14	5937,31	4108,17
HF2	1780,83	5644,78	3863,95
HF3	1768,29	5461,14	3692,85

Selleks, et kontrollida tabelis 4 esitatud Markovi mudeli suutlikkust sisendandmeid kirjeldada, tekitasime selle põhjal 10000 sünteetilise patsiendiga andmestiku. Genereeritud andmete valideerimiseks kasutasime astaktesti. Markovi mudeli järgi genereeritud trajektorid olid sarnased reaalsele vaatlusandmetele. Esiteks võib näha (vt lisa 14), et erinevate seisundite osakaalud olid nihkes enamasti vaid protsendi võrra. Teiseks suutis

Markovi mudel hästi kirjeldada peaaegu kõiki üleminekuid. Vaatleme ainult seisundite HF0, HF1, HF2, HF3, HFD vahelisi liikumisi. Kui rakendada väljastatud p-väärtustele Bonferroni korrigeerimist ($0,05/25 = 0,002$), siis ei ole astaktesti järgi ainult sarnased liikumised HF0'st HF1'te, HF2'te ja HF3'e. Joonisel 10 on kujutatud sünteetiliste patsientide seisundite jaotusi nii tavalise ravi kui telemeetriaga ravi korral. Jooniselt paistab silma, et genereeritud andmete puhul ei sure kõik patsiendid vaatlusperioodi 60. kuuks (samuti ka Suurbritannia uuringus). See tuleneb sellest, et Markovi mudelisse ei saa üldiselt sisse kirjutada, et see peab võtma mingiks ajahetkeks mingi kindla seisundi väärtuse. Teiste sõnadega võib öelda, et valminud Markovi mudel alahindab patsientide suremise intensiivsust.

6 Kokkuvõte

Antud magistritöö eesmärgiks oli välja arendada kaks R-paketti, mis töötavad OHDSI loodud lahendustel ning aitavad kaasa patsientide ravitrajektooride uurimisele. Töö esimeses osas tutvustasime lugejale Markovi ahelate olemust, tervishoiuökonomikat ja OHDSI kogukonda ning selle missiooni. Töö teises osas andsime ülevaate kahest R-paketist, mis käesoleva magistritöö käigus valmisid. Esimese paketi *Cohort2Trajectory* eesmärk on tekitada patsiendi meditsiinilisele ajaloole toetudes defineeritud seisunditest patsientide lineaarseid ravitrajektoore. Teise paketi *TrajectoryMarkovAnalysis* eesmärk on modelleerida patsientide ravitrajektoore Markovi mudelitega ning kirjeldada tervise seisundite maksumusi. Töö viimases osas viisime läbi nädisuuringu eelmainitud pakettide funktsionaalsuste demonstreerimiseks. Selleks koostasime Eesti südamepuudulikkusega inimeste ravitrajektooridel Markovi mudeli ning võrdlesime seda Suurbritannia andmetel saadud leidudega. Kasutades loodud pakette õnnestus uuringut reprodutseerida. Mõlema uuringu põhjal koostatud mudelid osutusid sarnasteks. Lisaks leidsime analoogselt Suurbritannia uuringuga, et telemeetria kasutamine südamepuudulikkusega patsientidel annab küll paremaid tulemusi, kuid maksab liiga palju.

Töö tulemusena valminud R-paketid on kasutatavad kõikvõimalike meditsiiniliste uurin-gute tegemiseks OHDSI OMOP ühtset andmemudelit kasutatavatel andmebaasidel. Töö-voole tulemusena väljastatud patsientide trajektoorigid on kasutatavad mitmesuguste ana-lüüside tegemiseks, antud töös tõime välja modelleerimise Markovi mudelitega ning kulutõhususanalüüsi tegemise. Paketid jälgivad OHDSI pakettide parimat praktikat, li-saks on neid mugav rakendada, kuna pakettidest on kasutada nii kasutajaliidesega kui käsurea versioon. Lisaks on pakettidel tehtud uuringud reprodutseeritavad ja jagatavad. Kuigi mõlemas pakettis on juba palju funktsioone, leidub erinevaid aspekte, mida saaks veel arendada või lisada. Magistritöö käigus valminud R-pakette on plaanis minna tut-vustama sellel aastal toimuvale Euroopa OHDSI sümposiumile, kus avaneb võimalus pakette OHDSI kogukonnale reklaamida ning leida koostööpartnereid.

Viidatud kirjandus

- [Chang et al., 2021] Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A. ja Borges, B. (2021). Package ‘shiny’: Web Application Framework for R. <https://CRAN.R-project.org/package=shiny> (17.05.2022).
- [Eddelbuettel ja Balamuta, 2018] Eddelbuettel, D. ja Balamuta, J. (2018). Extending R with C++: A Brief Introduction to Rcpp. *The American Statistician*, 72(1):28–36. <https://doi.org/10.1080/00031305.2017.1375990> (17.05.2022).
- [European Commission, 2022] European Commission (2022). Proposal for a Regulation - The European Health Data Space. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0197> (2022.05.17).
- [Giannoula et al., 2020] Giannoula, A., Centeno, E., Mayer, M., Sanz, F. ja Furlong, L. (2020). A System-Level Analysis of Patient Disease Trajectories Based on Clinical, Phenotypic and Molecular Similarities. *Bioinformatics*, 37(10):1435–1443. <https://doi.org/10.1093/bioinformatics/btaa964> (17.05.2022).
- [Hamra et al., 2013] Hamra, G., MacLehose, R. ja Richardson, D. (2013). Markov Chain Monte Carlo: An Introduction for Epidemiologists. *International Journal of Epidemiology*, 42(2):627–634. <https://doi.org/10.1093/ije/dyt043> (17.05.2022).
- [Hripesak et al., 2015] Hripesak, G. et al. (2015). Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. 216:574–578.
- [Hripesak et al., 2016] Hripesak, G., Ryan, P., Duke, J., Shah, N., Park, R., Huser, V., Suchard, M., Schuemie, M., DeFalco, F., Perotte, A., Banda, J., Reich, C., Schilling, L., Matheny, M., Meeker, D., Pratt, N. ja Madigan, D. (2016). Characterizing Treatment Pathways at Scale Using the OHDSI Network. *Proceedings of the National Academy of Sciences*, 113(27):7329–7336. <https://doi.org/10.1073/pnas.1510502113> (17.05.2022).
- [Hripesak et al., 2021] Hripesak, G., Schuemie, M., Madigan, D., Ryan, P. ja Suchard, M. (2021). Drawing Reproducible Conclusions From Observational Clinical Data With OHDSI.
- [Hu et al., 2019] Hu, Z., Tylavsky, F., Han, J. et al. (2019). Maternal Metabolic Factors During Pregnancy Predict Early Childhood Growth Trajectories and Obesity Risk: the CANDLER Study. *Int J Obes*, (43):1914–1922. <https://doi.org/10.1038/s41366-019-0326-z> (17.05.2022).

- [Jackson, 2011] Jackson, C. (2011). Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software*, 38(8):1–29. <https://doi.org/10.18637/jss.v038.i08> (17.05.2022).
- [Jensen et al., 2014] Jensen, A., Moseley, P., Oprea, T., Ellesøe, S., Eriksson, R., Schmock, H., Jensen, P., Jensen, L. ja Brunak, S. (2014). Temporal Disease Trajectories Condensed From Population-Wide Registry Data Covering 6.2 Million Patients. <https://doi.org/10.1038/ncomms5022> (17.05.2022).
- [Kassambara et al., 2021] Kassambara, A. et al. (2021). Package 'survminer': Drawing Survival Curves Using 'ggplot2'. <https://github.com/kassambara/survminer> (17.05.2022).
- [Kent et al., 2021] Kent, S., Burn, E., Dawoud, D. et al. (2021). Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment. *Pharmacoeconomics*, 39:275–285. <https://doi.org/10.1007/s40273-020-00981-9> (17.05.2022).
- [Kobelt, 2013] Kobelt, G. (2013). Health Economics: An Introduction to Economic Evaluation. Third Edition.
- [Komorowski ja Raffa, 2016] Komorowski, M. ja Raffa, J. (2016). Markov Models and Cost Effectiveness Analysis: Applications in Medical Research. *Secondary Analysis of Electronic Health Records*. https://doi.org/10.1007/978-3-319-43742-2_24 (17.05.2022).
- [Käärik, 2017] Käärik, M. (2017). Juhuslikud protsessid (MTMS.02.003).
- [Künnapuu et al., 2022] Künnapuu, K., Ioannou, S., Ligi, K., Kolde, R., Laur, S., Vilo, J., Rijnbeek, P. ja 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. <https://doi.org/10.1093/jamiaopen/ooac021> (17.05.2022).
- [Künnapuu et al., 2021] Künnapuu, K., Ligi, K., Kolde, R., Reisberg, S. ja Laur, S. (2021). Package 'Trajectories': Tool to Detect and Visualize Temporal Event Trajectories in OMOP Common Data Model. <https://github.com/EHDEN/Trajectories> (17.05.2022).
- [Ma et al., 2021] Ma, S., Dang, D., Wang, W., Wang, Y. ja Liu, L. (2021). Concentration Optimization of Combinatorial Drugs Using Markov Chain-Based Models. <https://doi.org/10.1186/s12859-021-04364-5> (17.05.2022).

- [Markus, 2021] Markus, A. (2021). Package 'TreatmentPatterns': A Package to Analyse Real-World Treatment Patterns of a Study Population of Interest. <https://github.com/mi-erasmusmc/TreatmentPatterns> (17.05.2022).
- [Menachemi ja Collum, 2011] Menachemi, N. ja Collum, T. (2011). Benefits and Drawbacks of Electronic Health Record Systems. 4:47–55. <https://doi.org/10.2147/RMHP.S12985> (17.05.2022).
- [Norris, 1998] Norris, J. (1998). Markov Chains.
- [OHDSI, 2021] OHDSI (2021). The Book of OHDSI: Observational Health Data Sciences and Informatics. <https://ohdsi.github.io/TheBookOfOhdsi/>(17.05.2022).
- [Pajusalu, 2022] Pajusalu, M. (2022). Package 'TrajectoryViz': Visualize Trajectories (Drug Sequence, State Sequence) and Extract Detailed Paths. <https://github.com/HealthInformaticsUT/TrajectoryViz> (17.05.2022).
- [Pandor et al., 2013] Pandor, A. et al. (2013). Home Telemonitoring or Structured Telephone Support Programmes After Recent Discharge in Patients With Heart Failure: Systematic Review and Economic Evaluation. *Health technology assessment (Winc- hester, England)*, 17:1–208. <https://doi.org/10.3310/hta17320> (17.05.2022).
- [Pokrovski et al., 2019] Pokrovski, M. et al. (2019). Tervishoiuteenuste loetelu komisjo- ni koosoleku kokkuvõte. *Eesti Haigekassa*. <https://tinyurl.com/mr2ejsyz> (17.05.2022).
- [Reps et al., 2018] Reps, J., Schuemie, M., Suchard, M., Ryan, P. ja Rijnbeek, P. (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> (17.05.2022).
- [Romeu, 2020] Romeu, J. (2020). A Markov Chain Model for Covid-19 Survival Analysis. <https://doi.org/10.13140/RG.2.2.36349.18408> (17.05.2022).
- [Schoenfeld, 1981] Schoenfeld, D. (1981). The Asymptotic Properties of Nonpara- metric Tests for Comparing Survival Distributions. *Biometrika*, 68:316–319. <https://doi.org/10.1093/biomet/68.1.316> (17.05.2022).
- [Schuemie, 2020] Schuemie, M. (2020). Package 'Eunomia': A Standard Dataset in the OMOP Common Data Model. <https://github.com/OHDSI/Eunomia> (17.05.2022).
- [Schuemie et al., 2019] Schuemie, M., Ryan, P., Man, K., Wong, I., Suchard, M. ja Hripcsak, G. (2019). A Plea to Stop Using the Case-Control Design in Retrospective Database Studies. 38:4199– 4208. <https://doi.org/10.1002/sim.8215> (17.05.2022).

- [Schuemie ja Suchard, 2022a] Schuemie, M. ja Suchard, M. (2022a). DatabaseConnector: Connecting to Various Database Platforms. <https://ohdsi.github.io/DatabaseConnector/> (17.05.2022).
- [Schuemie ja Suchard, 2022b] Schuemie, M. ja Suchard, M. (2022b). Package 'SqlRender': Rendering Parameterized SQL and Translation to Dialects. <https://github.com/OHDSI/SqlRender> (17.05.2022).
- [Soper et al., 2020] Soper, B., Nygård, M., Abdulla, G., Meng, R. ja Nygård, J. (2020). A Hidden Markov Model for Population-Level Cervical Cancer Screening Data. <https://doi.org/10.1002/sim.8681> (17.05.2022).
- [Spruance et al., 2004] Spruance, S., Reid, J., Grace, M. ja Samore, M. (2004). Hazard Ratio in Clinical Trials. *Antimicrob Agents Chemothe*, 48(8):2787–2792. <https://doi.org/10.1128/AAC.48.8.2787-2792.2004> (17.05.2022).
- [Thokala et al., 2020] Thokala, P., Dodd, P., Baalbaki, H., Brennan, A., Dixon, S. ja Lowrie, K. (2020). Developing Markov Models From Real-World Data: A Case Study of Heart Failure Modeling Using Administrative Data. *Value in Health*, 23(6):743–750. <https://doi.org/10.1016/j.jval.2020.02.012> (17.05.2022).
- [Tian et al., 2018] Tian, Y., Schuemie, M. ja Suchard, M. (2018). Evaluating Large-Scale Propensity Score Performance Through Real-World and Synthetic Data Experiments. pages 2005–2014. <https://doi.org/10.1093/ije/dyy120> (17.05.2022).
- [Vabariigi Valitsus, 2022] Vabariigi Valitsus (2022). Eesti Haigekassa tervishoiuteenuste loetelu. *Riigi Teataja*.
- [Voss et al., 2017] Voss, E. et al. (2017). Large-Scale Adverse Effects Related to Treatment Evidence Standardization (LAERTES): An Open Scalable System for Linking Pharmacovigilance Evidence Sources With Clinical Data. <https://doi.org/10.1186/s13326-017-0115-3> (17.05.2022).

1 Paketi *Cohort2Trajectory* kasutajaliidese vaheleht "Importimine ATLAS'st"

Choose the patient cohort:

- 149
- 150
- 151
- 152
- 155
- 156
- 157
- 158
- 159
- 160
- 161
- 162
- 163
- 164
- 165
- 166
- 167
- 168
- 169
- 170
- 172
- 173
- 174
- 175
- 176

Choose the state cohorts:




- 149
- 150
- 151
- 152
- 155
- 156
- 157
- 158
- 159
- 160
- 161
- 162
- 163
- 164
- 165
- 166
- 167
- 168
- 169
- 170
- 172
- 173
- 174
- 175
- 176

Import data

Selected cohorts Customise Cohorts' overview

Information

You have chosen target cohort 155 and state cohorts 151, 156, 158, 159, 160

 SAMPLE 3224	 UNIQUE 3359	 REPRESENTATIVE 100 %
---	---	--

Selected cohorts Customise Cohorts' overview

Name the selected patient cohort

Name the selected state cohorts

Apply

46

2 Paketi *Cohort2Trajectory* kasutajaliidese vaheleht "Importimine JSON'st"




Patient cohort JSON:

Target cohort JSON

```
{
  "cdmVersionRange": ">=5.0.0",
  "PrimaryCriteria": {
    "CriteriaList": [
      {
        "VisitOccurrence": {
          "VisitTypeExclude": false
        }
      }
    ]
  },
  "ObservationWindow": {
    "PriorDays": 0,
    "PostDays": 0
  },
  "PrimaryCriteriaLimit": 1
}
```

Show saved JSONs Add state Remove state Import data

Information

 SAMPLE 3224	 UNIQUE 3359	 REPRESENTED 100 %
---	---	---

State cohort 1 JSON:

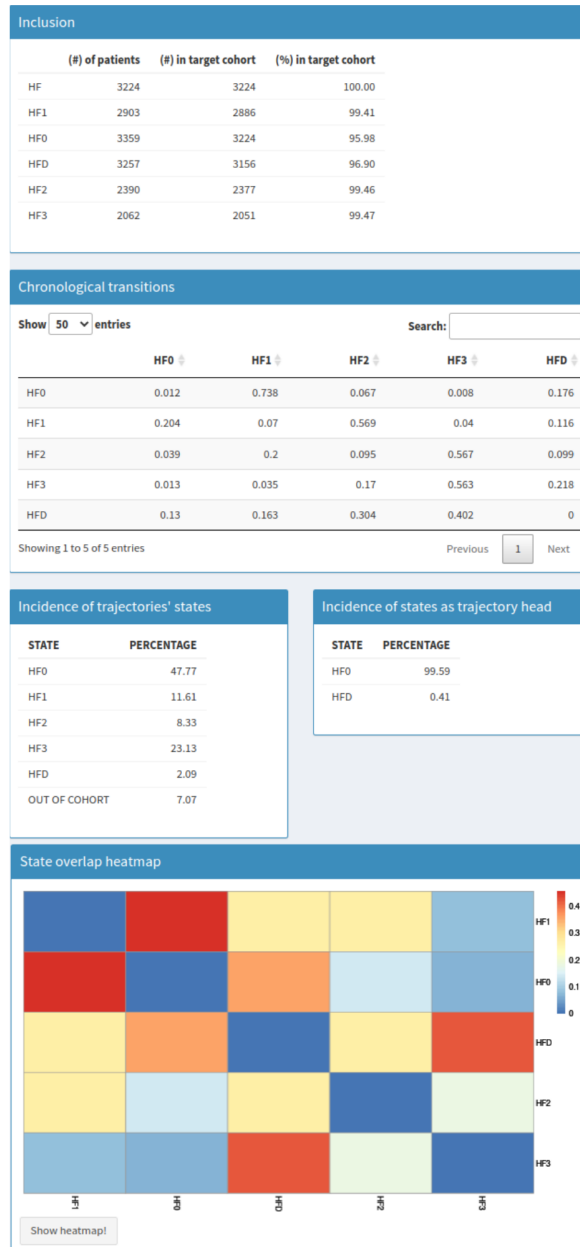
State cohort label

HF0

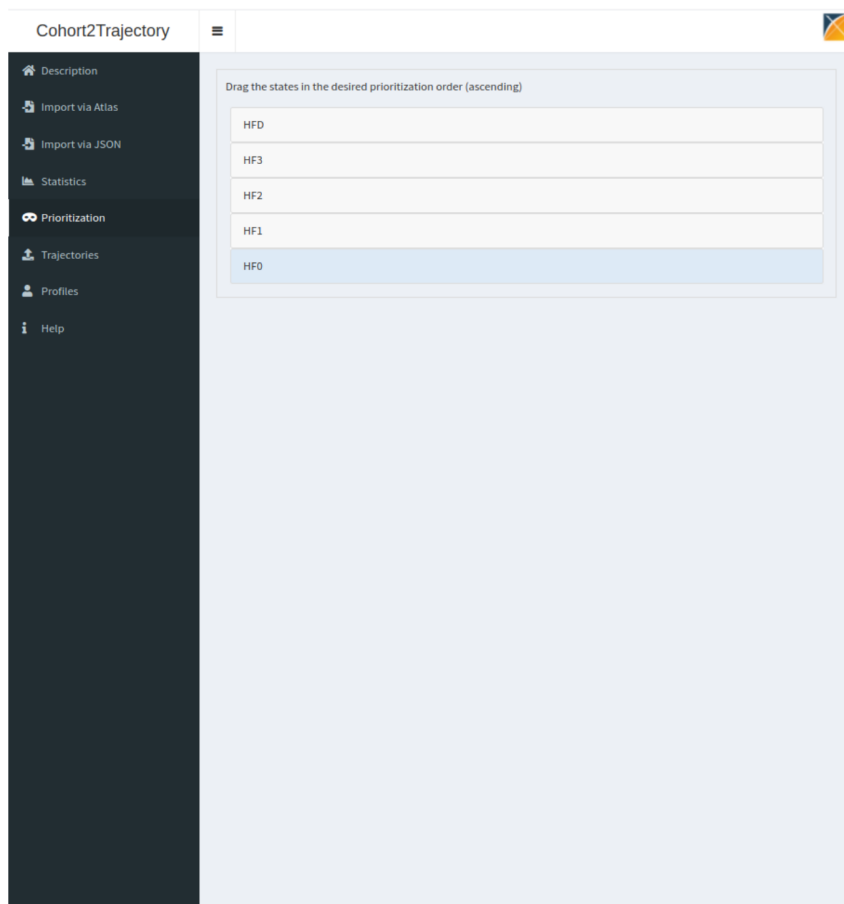
State cohort JSON

```
{
  "cdmVersionRange": ">=5.0.0",
  "PrimaryCriteria": {
    "CriteriaList": [
      {
        "VisitOccurrence": {
          "VisitTypeExclude": false
        }
      }
    ]
  },
  "ObservationWindow": {
    "PriorDays": 0,
    "PostDays": 0
  },
  "PrimaryCriteriaLimit": 1
}
```

3 Paketi *Cohort2Trajectory* kasutajaliidese vaheleht "Statistika"



4 Paketi *Cohort2Trajectory* kasutajaliidese vaheleht "Prioritiseerimine"



5 Paketi *Cohort2Trajectory* kasutajaliidese vaheleht "Trajektoorid"

Trajectory generation settings

Trajectory type: <input checked="" type="radio"/> Discrete time <input type="radio"/> Continuous time	Length of a state stay in days (only discrete case) <input type="text" value="30"/>	State selection type <input type="radio"/> First occurring <input type="radio"/> Largest overlap <input checked="" type="radio"/> Priority first
Choose absorbing states: <input type="checkbox"/> No absorbing state <input type="checkbox"/> HF0 <input type="checkbox"/> HF1 <input type="checkbox"/> HF2 <input type="checkbox"/> HF3 <input checked="" type="checkbox"/> HFD	Choose mandatory states: <input type="checkbox"/> No mandatory state <input type="checkbox"/> HF0 <input type="checkbox"/> HF1 <input type="checkbox"/> HF2 <input type="checkbox"/> HF3 <input checked="" type="checkbox"/> HFD	Allow states out of observation period <input type="radio"/> No <input checked="" type="radio"/> Yes <input type="button" value="Generate"/>

Generating trajectories

6 Paketi *Cohort2Trajectory* kasutamine: ATLAS-tarkvaras kohordi defineerimine

Defineerime OMOP CDM kasutades kohordi, kuhu kuuluvad kõik patsiendid, kellel on registreeritud vähemalt üks hospitaliseerimine südamepuudulikkuse tõttu viis aastat enne nende surma. Selleks tuleb meil ATLAS-tarkvaras avada vaheleht ”*Cohort Definitions*” ning täita ära kohordi moodustamiseks vajalike tingimuste ankeet. Esimese asjana tuleb meil defineerida kohorti sisenemise tingimus (vt joonis 1). Selleks võib olla mõne haiguse esinemine, surm, mõne seadme kasutamine, ravimite manustamise periood, mõõtmine, täheldatud nähtus, kliinilise protseduuri läbiviimine jms. Kuna meie seisukohast on oluline saada patsiendi vaatlusperioodi alguseks kuupäev viis aastat enne tema surma, siis võtame iga patsiendi kohorti tema varaseima visiidi hetkel, mis jääb viie aasta perioodi sisse, mille jooksul patsient sureb.

Cohort #155

Heart failure with hospitalizations [Target cohort]

Definition | Concept Sets | Generation | Samples | Reporting | Export | Versions | Messages

Enter a cohort definition description here

Cohort Entry Events

Events having any of the following criteria:

+ Add Initial Event...

a visit occurrence of Any Visit + Add attribute... Delete Criteria

with continuous observation of at least 0 days before and 0 days after event index date

Limit initial events to: earliest event per person.

Restrict initial events to:

having all of the following criteria: + Add criteria to group...

with at least 1 using all occurrences of:

a death occurrence from Any Death + Add attribute... Delete Criteria

where event starts between 0 days Before and 1827 days After index start date [add additional constraint](#)

The index date refers to the event from the Cohort Entry criteria.

allow events from outside observation period

Limit initial events to: earliest event per person.

Remove initial event restriction

Joonis 1. Kohorti sisenemise tingimuste defineerimine

Teiseks tuleb lisada kohordi definitsioonile kaasa arvamise kriteeriumid. Neid oli meie uuringu sihtkohordil üks: patsiendil pidi esinema vähemalt üks südamepuudulikkusega seotud hospitaliseerimine. Selle kehtestamiseks lisame tingimuseks, et patsiendil peab esinema vähemalt üks haiglaviisit, millega on seotud mõni südamepuudulikkuse diagnoos.

The screenshot shows the 'Inclusion Criteria' configuration interface. At the top, there is a header 'Inclusion Criteria' with a help icon. Below it, a green button 'New inclusion criteria' is visible. The main title is 'Heart Failure', with 'Copy' and 'Delete' buttons. A list on the left shows '1. Heart Failure'. The main area contains a text input 'enter an inclusion rule description' and a dropdown 'having all of the following criteria:'. Below this is a nested structure of criteria:

- Outer criteria: 'with at least 1 using all occurrences of: a visit occurrence of Outpatient & Inpatient visits'. It includes a '+ Add attribute...' button and a 'Delete Criteria' button.
- Inner criteria (marked with a red X): 'having all of the following criteria:'.
 - Inner criteria: 'with at least 1 using all occurrences of: a condition occurrence of HF'. It includes a '+ Add attribute...' button and a 'Delete Criteria' button.
 - Temporal constraints: 'where event starts between All days Before and All days After'. It includes an 'index start date' button and a link 'add additional constraint'. A note states: 'The index date refers to the visit of Outpatient & Inpatient visits.' There are checkboxes for 'restrict to the same visit occurrence' (checked) and 'allow events from outside observation period'.
- Bottom criteria: 'where event starts between All days Before and All days After'. It includes an 'index start date' button and a link 'add additional constraint'. A note states: 'The index date refers to the event from the Cohort Entry criteria.' There are checkboxes for 'restrict to the same visit occurrence' and 'allow events from outside observation period'.

At the bottom, there is a field 'Limit qualifying events to: earliest event per person.'

Joonis 2. Kohorti kuulumise kriteeriumite defineerimine

Lõpetuseks tuleb määrata kohordist väljumise kriteerium. Antud juhul on selleks patsiendi surm. Lisades selle olemegi defineerinud patsientide kohordi, kus on vaatlusperioodid kõikidest andmebaasis olevatest patsientidest, kellel on esinenud vähemalt üks südamepuudulikkusega seotud hospitaliseerimine ning kes surevad vaatlusperioodi lõpuks.

Cohort Exit ?

Event Persistence:
 Event will persist until: fixed duration relative to initial event ▼

Fixed Duration Persistence:
 The event end date is derived from adding a number of days to the event's start or end date. If an offset is added to the event's start date, all cohort episodes will have the same fixed duration (subject to further censoring). If an offset is added to the event's end date, persons in the cohort may have varying cohort duration times due to the varying event durations (such as eras of persistent drug exposure or visit length of stay). This event persistence assures that the cohort end date will be no greater than the selected index event date, plus the days offset.

- Event date to offset from: start date ▼
- Number of days offset: 1827 ▼ days

Censoring Events:
 Exit Cohort based on the following criteria:

a death occurrence from

Any Death

▼

+ Add Censoring Event... ▼

+ Add attribute... ▼

Delete Criteria

Joonis 3. Kohordist väljumise kriteeriumite defineerimine

7 Paketi Cohort2Trajectory kasutamine: näide väljastatud trajektoorie failist

```
SUBJECT_ID STATE STATE_START_DATE STATE_END_DATE TIME_IN_COHORT GENDER_CONCEPT_ID AGE STATE_ID
244 START 2017-03-16 2017-03-16 0 8532 85.372 1
244 HF0 2017-03-16 2017-04-15 0 8532 85.372 2
244 HF0 2017-04-15 2017-05-15 0.082135523613963 8532 85.454 2
244 HF1 2017-05-15 2017-06-14 0.164271047227926 8532 85.536 5
244 HF1 2017-06-14 2017-07-14 0.246406570841889 8532 85.618 5
244 HF2 2017-07-14 2017-08-13 0.328542094455852 8532 85.7 7
244 HF2 2017-08-13 2017-09-12 0.410677618069815 8532 85.782 7
244 HF3 2017-09-12 2017-10-12 0.492813141683778 8532 85.864 8
244 HF3 2017-10-12 2017-11-11 0.574948665297741 8532 85.947 8
244 HF3 2017-11-11 2017-12-11 0.657084188911704 8532 86.029 8
244 HF3 2017-12-11 2018-01-10 0.739219712525667 8532 86.111 8
244 HF3 2018-01-10 2018-02-09 0.82135523613963 8532 86.193 8
244 HF3 2018-02-09 2018-03-11 0.903490759753593 8532 86.275 8
244 HF3 2018-03-11 2018-04-10 0.985626283367556 8532 86.357 8
244 HF3 2018-04-10 2018-05-10 1.06776180698152 8532 86.439 8
244 HF3 2018-05-10 2018-06-09 1.14989733059548 8532 86.522 8
244 HF3 2018-06-09 2018-07-09 1.23203285420945 8532 86.604 8
244 HF3 2018-07-09 2018-08-08 1.31416837782341 8532 86.686 8
244 HF3 2018-08-08 2018-09-07 1.39630390143737 8532 86.768 8
244 HF3 2018-09-07 2018-10-07 1.47843942505133 8532 86.85 8
244 HF3 2018-10-07 2018-11-06 1.5605749486653 8532 86.932 8
244 HF3 2018-11-06 2018-12-06 1.64271047227926 8532 87.014 8
244 HF3 2018-12-06 2019-01-05 1.72484599589322 8532 87.097 8
244 HF3 2019-01-05 2019-02-04 1.80698151950719 8532 87.179 8
244 HF3 2019-02-04 2019-03-06 1.88911704312115 8532 87.261 8
244 HF3 2019-03-06 2019-04-05 1.97125256673511 8532 87.343 8
244 HF3 2019-04-05 2019-05-05 2.05338809034908 8532 87.425 8
244 HF3 2019-05-05 2019-06-04 2.13552361396304 8532 87.507 8
244 HF3 2019-06-04 2019-07-04 2.217659137577 8532 87.589 8
244 HF3 2019-07-04 2019-08-03 2.29979466119096 8532 87.671 8
244 HF3 2019-08-03 2019-09-02 2.38193018480493 8532 87.754 8
244 HF3 2019-09-02 2019-10-02 2.46406570841889 8532 87.836 8
244 HF3 2019-10-02 2019-11-01 2.54620123203285 8532 87.918 8
244 HF3 2019-11-01 2019-12-01 2.62833675564682 8532 88 8
244 HF3 2019-12-01 2019-12-31 2.71047227926078 8532 88.082 8
244 HF3 2019-12-31 2020-01-30 2.79260780287474 8532 88.164 8
244 HF3 2020-01-30 2020-02-29 2.87474332648871 8532 88.246 8
244 HF3 2020-02-29 2020-03-30 2.95687885010267 8532 88.329 8
244 HF3 2020-03-30 2020-04-29 3.039014373171663 8532 88.411 8
244 HF3 2020-04-29 2020-05-29 3.1211498973306 8532 88.493 8
244 HF3 2020-05-29 2020-06-28 3.20328542094456 8532 88.575 8
244 HF3 2020-06-28 2020-07-28 3.28542094455852 8532 88.657 8
244 HF3 2020-07-28 2020-08-27 3.36755646817248 8532 88.739 8
244 HF3 2020-08-27 2020-09-26 3.44969199178645 8532 88.821 8
244 HF3 2020-09-26 2020-10-26 3.53182751540041 8532 88.903 8
244 HF3 2020-10-26 2020-11-25 3.61396303901437 8532 88.986 8
244 HF3 2020-11-25 2020-12-25 3.69609856262834 8532 89.068 8
244 HF3 2020-12-25 2021-01-24 3.7782340862423 8532 89.15 8
244 HF3 2021-01-24 2021-02-23 3.86036960985626 8532 89.232 8
244 HF3 2021-02-23 2021-03-25 3.94250513347023 8532 89.314 8
244 HF3 2021-03-25 2021-04-24 4.02464065708419 8532 89.396 8
244 HF3 2021-04-24 2021-05-24 4.10677618069815 8532 89.478 8
244 HF2 2021-05-24 2021-06-23 4.18891170431212 8532 89.561 7
244 HFD 2021-06-23 2021-07-23 4.27104722792608 8532 89.643 3
244 EXIT 2021-07-24 2021-07-24 4.27104722792608 8532 89.728 4
315 START 2017-05-12 2017-05-12 0 8507 83.113 1
315 HF0 2017-05-12 2017-06-11 0 8507 83.113 2
315 HF0 2017-06-11 2017-07-11 0.082135523613963 8507 83.195 2
315 HF0 2017-07-11 2017-08-10 0.164271047227926 8507 83.277 2
315 HF0 2017-08-10 2017-09-09 0.246406570841889 8507 83.359 2
315 HF0 2017-09-09 2017-10-09 0.328542094455852 8507 83.441 2
315 HF0 2017-10-09 2017-11-08 0.410677618069815 8507 83.524 2
315 HF0 2017-11-08 2017-12-08 0.492813141683778 8507 83.606 2
315 HF0 2017-12-08 2018-01-07 0.574948665297741 8507 83.688 2
315 HFD 2018-01-07 2018-02-07 3.61396303901437 8507 83.770 3
315 EXIT 2018-02-08 2018-02-08 3.61396303901437 8507 83.773 4
```

8 Paketi *TrajectoryMarkovAnalysis* kasutajaliidese vaheleht "Importimine"

Import the patient treatment trajectory data:

Choose CSV File (MAX 150MB)

Browse... HF_FINAL2.csv

Upload complete

Preview of the imported dataset:

SUBJECT_ID	STATE	STATE_START_DATE	STATE_END_DATE
115	START	2017-05-12	2017-05-12
115	HF0	2017-05-12	2017-06-11
115	HF0	2017-06-11	2017-07-11
115	HF0	2017-07-11	2017-08-10
115	HF0	2017-08-10	2017-09-09

SunburstPlot of the imported data:

HF0 HF1 HF2 HF3 HF4 10.1%

Legend

- HF0
- HF1
- HF2
- HF3
- End

10.1%
320 of 3.2k

9 Paketi *TrajectoryMarkovAnalysis* kasutajaliidese vaheleht "Diskreetse ajaga Markovi ahelad"

Initial settings

Select the cost domains included in analysis:

- Drug
- Visit
- Procedure
- Device
- Measurement
- Observation
- Specimen

Select the states you want to exclude from analysis:

- START
- HF0
- HFD
- EXIT
- HF1
- OUT OF COHORT
- HF2
- HF3

Stochastic transitions

Show entries
Search:

	START	HF0	HFD	HF1	HF2	HF3	EXIT
START	0	0.9959	0.0041	0	0	0	0
HF0	0	0.9389	0.0146	0.0371	0.0075	0.0019	0
HFD	0	0	0	0	0	0	1
HF1	0	0.056	0.0238	0.7807	0.1167	0.0228	0
HF2	0	0.015	0.0294	0.0724	0.7026	0.1805	0
HF3	0	0.0026	0.0373	0.0073	0.0331	0.9197	0
EXIT	0	0	0	0	0	0	1

Showing 1 to 7 of 7 entries
Previous Next

Trajectory start statistics

Show entries

Search:

	STATE	PERCENTAGE	MEAN CHARGE	MEAN COST	MEAN PAID
1	HF0	99.59%	€9,318.03	€9,318.03	€0.00
2	HFD	0.41%	€815.57	€815.57	€0.00

Showing 1 to 2 of 2 entries

Previous Next

The mean trajectory cost of a patient: 9505 €
The median trajectory cost of a patient: 7375 €

State statistics (cost per 1 day)

Show entries

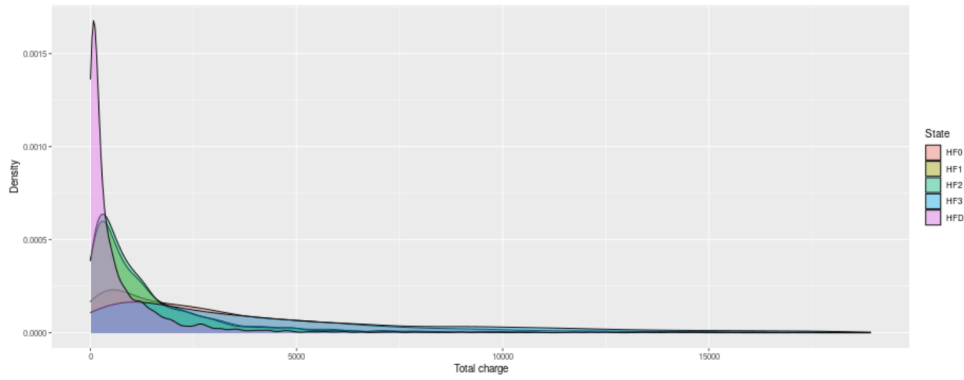
Search:

	STATE	PERCENTAGE	MEAN CHARGE	CI CHARGE	MEAN COST	CI COST	MEAN PAID	CI PAID
1	HF0	51.40%	€8.11	(0, 49.11)	€8.11	(0, 49.11)	€0.00	(0, 0)
2	HF1	12.49%	€8.87	(0, 47.86)	€8.87	(0, 47.86)	€0.00	(0, 0)
3	HF2	8.96%	€9.17	(0, 54.18)	€9.17	(0, 54.18)	€0.00	(0, 0)
4	HF3	24.89%	€10.51	(0, 51.53)	€10.51	(0, 51.53)	€0.00	(0, 0)
5	HFD	2.25%	€23.31	(0, 112.25)	€23.31	(0, 112.25)	€0.00	(0, 0)

Showing 1 to 5 of 5 entries

Previous Next

State cost distributions



Probability of transitions in x transfers

Number of transfers:

Calculate

Show 10 entries

Search:

	START	HF0	HFD	HF1	HF2	HF3	EXIT
START	0	0.742713358308059	0.0155079565991094	0.108485710144309	0.0423068292573506	0.0271333205877345	0.0638443554046864
HF0	0	0.707012672914857	0.0157460272826798	0.115986090151033	0.049054647764655	0.0366256925369683	0.0755621166821928
HFD	0	0	0	0	0	0	1
HF1	0	0.182171478542919	0.0228910718963947	0.2887775725792	0.184236091045991	0.204253612895578	0.117581630689548
HF2	0	0.0704305058776549	0.0269300126801447	0.121937206010045	0.195018916104806	0.441621725238142	0.14374964838407
HF3	0	0.0207773249995203	0.0294842462609861	0.0378514574552483	0.084772671094945	0.657360832147481	0.169721973946875
EXIT	0	0	0	0	0	0	1

Showing 1 to 7 of 7 entries

Previous **1** Next

10 Paketi *TrajectoryMarkovAnalysis* kasutajaliidese vaheleht "Pideva ajaga Markovi ahelad"

Initial intensity matrix Markov model

	START	HFD	HF3	HF2	HF1	HFO	EXIT
START	0	1	1	1	1	1	1
HFD	0	0	0	0	0	0	1
HF3	0	1	0	1	1	1	1
HF2	0	1	1	0	1	1	1
HF1	0	1	1	1	0	1	1
HFO	0	1	1	1	1	0	1
EXIT	0	0	0	0	0	0	0

Select the cost domains included in analysis:

- Drug
- Visit
- Procedure
- Device
- Measurement
- Observation
- Specimen

Select the states you want to exclude from analysis:

- START
- HFO
- HFD
- EXIT
- HF1
- HF2
- HF3

Calculated CMC intensity matrix

Show 10 entries Search:

	START	HFD	HF3	HF2	HF1	HFO	EXIT
START	-41.0713710867494	0.0390412272687732	0	0	0	41.0323298594806	0
HFD	0	-60.5209398509947	0	0	0	0	60.5209398509947
HF3	0	0.443025319303157	-0.88636306830963	0.345234822165013	0.0712339723562199	0.0268689544852408	0
HF2	0	0.352839311870489	2.01721948694378	-3.22662265460513	0.714963868790202	0.141599987000657	0
HF1	0	0.341050939883638	0.120226590318692	1.6923732790269	-2.75907756761979	0.605426758390558	0
HFO	0	0.117672589897709	0.00553370408924217	0.0498033368031795	0.54458158478248	-0.71759121557261	0
EXIT	0	0	0	0	0	0	0

Showing 1 to 7 of 7 entries Previous 1 Next

Calculated CMC sojourn time

Show 10 entries Search:

	Estimate (days)	Standard error	Lower	Upper	Estimated stay per patient (days)
START	8.8930559252218	0.15830057413229	8.58814238325299	9.20879512237039	8.89305592522183
HFD	6.03510125419833	0.107427638094203	5.82817754709117	6.24937158385037	6.03510138710589
HF3	412.077187169545	7.73657728504807	397.189371446799	427.523041633829	370.425532305379
HF2	113.198858093525	1.75296797567606	109.814720592863	116.687283858652	149.568833408927
HF1	132.381200255669	1.97386093815791	128.568486455838	136.306980537966	188.672572477876
HFO	508.994525119075	7.66554478628066	494.189905206308	524.242652210868	711.076318513015

Showing 1 to 6 of 6 entries Previous 1 Next

Calculated CMC transition matrix

Transition time in days:

Calculate

Show entries Search:

	START	HFD	HF3	HF2	HF1	HFO	EXIT
START	0.4551	0.0008	0	0.0003	0.0031	0.5401	0.0005
HFD	0	0.3135	0	0	0	0	0.6865
HF3	0	0.005	0.9833	0.0064	0.0014	0.0005	0.0035
HF2	0	0.004	0.0372	0.9404	0.013	0.0027	0.0028
HF1	0	0.0038	0.0028	0.0306	0.9488	0.0113	0.0027
HFO	0	0.0014	0.0001	0.0011	0.0101	0.9864	0.0009
EXIT	0	0	0	0	0	0	1

Showing 1 to 7 of 7 entries Previous Next

Trajectory start statistics

Show entries Search:

	STATE	PERCENTAGE	MEAN CHARGE	MEAN COST	MEAN PAID
1	HFO	99.90%	€9,297.17	€9,297.17	€0.00
2	HFD	0.10%	€200.20	€200.20	€0.00

Showing 1 to 2 of 2 entries Previous Next

State statistics (cost per 1 day)

Show entries

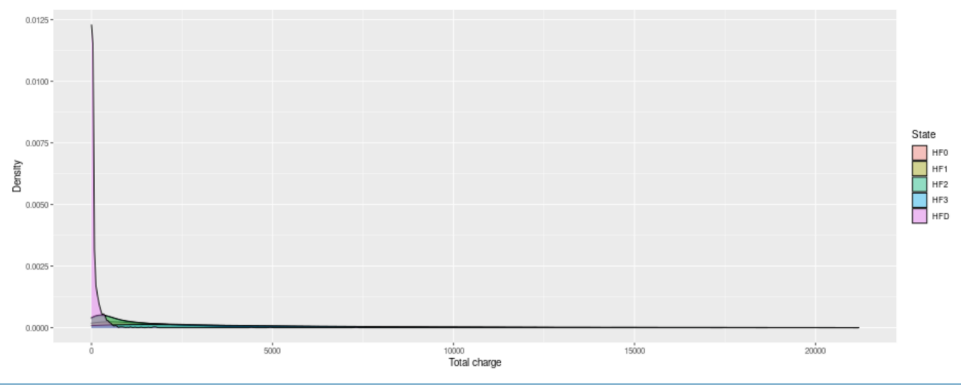
Search:

	STATE	PERCENTAGE	MEAN CHARGE	CI CHARGE	MEAN COST	CI COST	MEAN PAID	CI PAID
1	HF0	18.94%	€22.90	(0 , 140.1)	€22.90	(0 , 140.1)	€0.00	(0 , 0)
2	HF1	20.54%	€53.34	(0 , 461.05)	€53.34	(0 , 461.05)	€0.00	(0 , 0)
3	HF2	19.55%	€44.24	(0 , 232.63)	€44.24	(0 , 232.63)	€0.00	(0 , 0)
4	HF3	27.57%	€37.43	(0 , 330.78)	€37.43	(0 , 330.78)	€0.00	(0 , 0)
5	HFD	13.40%	€91.06	(0 , 643.12)	€91.06	(0 , 643.12)	€0.00	(0 , 0)

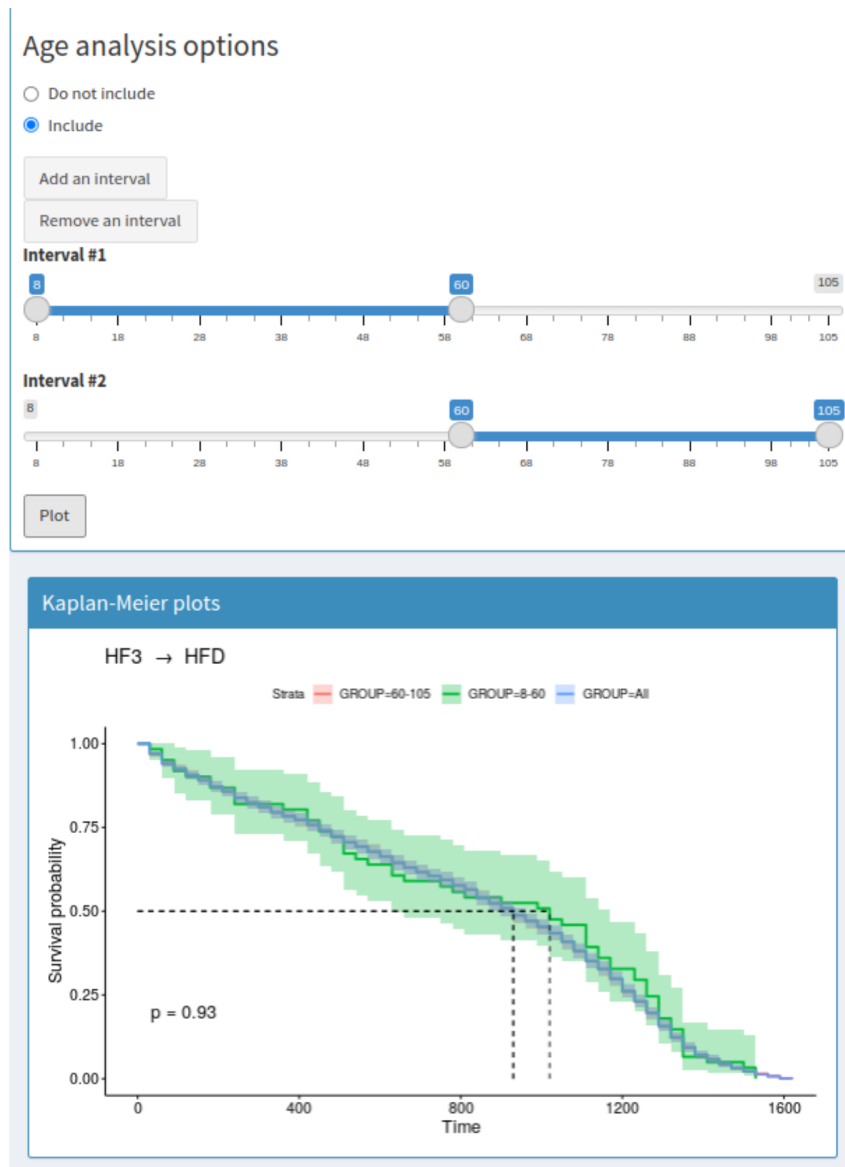
Showing 1 to 5 of 5 entries

Previous Next

State cost distributions



11 Paketi *TrajectoryMarkovAnalysis* kasutajaliidese vaheleht "Kaplan-Meieri graafikud"



12 Paketi *TrajectoryMarkovAnalysis* kasutajaliidese vaheleht "Otsustuspuud"

DT architecture

Add layer Remove layer Show tree structure

Layer #1

- START
- HF0
- HFD
- EXIT
- HF1
- HF2
- HF3

Layer #2

- START
- HF0
- HFD
- EXIT
- HF1
- HF2
- HF3

Layer #3

- START
- HF0
- HFD
- EXIT
- HF1
- HF2
- HF3

Import

Generate on Data

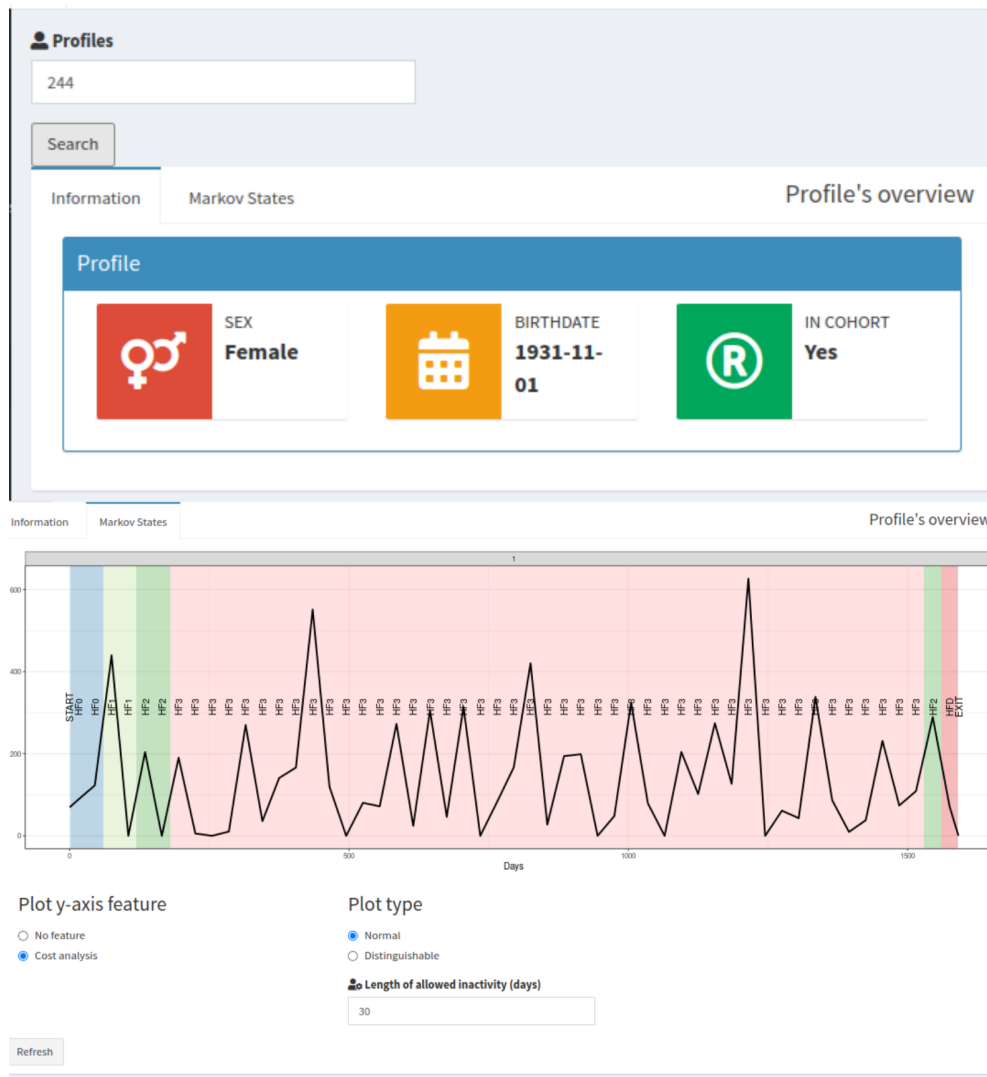
Select probability type

- Probability in target cohort
- Probability on layer level

Probability tree

The diagram illustrates a probability tree starting from a root node labeled "START". A horizontal line leads to a node labeled "HF0 100%". From "HF0", two curved lines branch out: an upper one to "HF2 16.78%" and a lower one to "HF1 83.22%". From "HF2", two curved lines branch out: an upper one to "HF3 86.02%" and a lower one to "HFD 13.98%". From "HF1", two curved lines branch out: an upper one to "HF3 49.08%" and a lower one to "HFD 50.92%".

13 Paketi *TrajectoryMarkovAnalysis* kasutajaliidese vaheleht "Profilid"



14 Paketi *TrajectoryMarkovAnalysis* kasutajaliidese vaheleht "Genereerimine"

Generation settings

Number of patients

Date range

2022-05-17	to	2022-09-01
------------	----	------------

Duration patient spends in a state

Maximum days out of cohort

Matrix to use: state selection type

Discrete time trajectory
 Continuous time trajectory

Generate cost per state from observed data?

No
 Yes

Observed vs generated: (%) of occurrence

STATE	OBSERVED	GENERATED
EXIT	2.0070%	2.1380%
HF0	45.8510%	49.4710%
HF1	11.1450%	12.0150%
HF2	7.9960%	8.3250%
HF3	22.2000%	23.7750%
HFD	2.0070%	2.1380%
START	2.0070%	2.1380%

Observed vs generated: LogRank tests for transferring between states

Show entries Search:

	EXIT	HF0	HF1	HF2	HF3	HFD	OUT OF COHORT	START
EXIT	1	1	1	1	1	1	1	1
HF0	1.86100657931495e-7	1	0.000345366190521526	4.78535026975653e-7	4.32221770276824e-12	0.0485319472283173	1	1
HF1	0.00100042725025348	1	1	0.82507604536335	0.141078067741474	0.053921136582359	1	1
HF2	0.000130105062969442	1	0.0234405980466247	1	0.758627769393603	0.0170644211179194	1	1
HF3	3.36470285755097e-75	1	0.0240572912129711	0.0142513035943813	1	0.0421267806960246	1	1
HFD	1	1	1	1	1	1	1	1
OUT OF COHORT	1	1	1	1	1	1	1	1
START	0.00123197198267004	1	7.33154214208224e-18	8.68154382809918e-27	2.64976840266055e-38	0.00012919514675161	1	1

Showing 1 to 8 of 8 entries

Previous Next

15 Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Markus Haug**,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose **Patsientide ravitrajektooride modelleerimine Markovi ahelatega**, mille juhendaja on Raivo Kolde, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Markus Haug
17.05.2022