

Rahvastikuandmete simulatsioonimudeli arendamine

Taotlus

2. september 2025

Tartu Ülikool

Eesti Rakendusuuringute Keskus CentAR

1. Üldandmed

1.1. Üldandmed

Uuringu nimi eesti keeles:

Rahvastikuandmete simulatsioonimudeli arendamine (Riikliku pensioni prognoosimudeli edasiarendamine ning kasutusvaldkonna laiendamine)

Uuringu nimi inglise keeles:

Development of a Population Simulation Model (Further development and Expansion of Estonian Pension Forecasting Model)

Uuringu algus: 1. november 2025

Uuringu lõpp: 29. oktoober 2027

Vastutav täitja: Tiit Tammaru

Juhtpartner: Tartu Ülikool

Partner(id): Eesti Rakendusuuringu Keskus CentAR OÜ (Centar)

Partnerite kinnituskirjad

Taotletav summa (ilma käibemaksuta): 341 200

Taotletav summa (koos käibemaksuga): 423 088

Valdkond ja eriala

Frascati Manuaali valdkond	ETIS valdkond	CERCS valdkond	%
5.7 Sotsiaal- ja majandusgeograafia	2. Ühiskonnateadused ja kultuur 2.11. Sotsiaalteadused	S250 Demograafia	100,0

Lühikokkuvõte

Eestis puudub mikrosimulatsioonimudel, mis võimaldaks koostada regionaalseid rahvastikuprognose ja suudaks arvesse võtta erinevate sotsiaaldemograafiliste tunnuste omavahelisi seoseid. Samas on Eestis hea kvaliteediga registriandmeid, mida saaks sellise mudeli arendamiseks ja rakendamiseks kasutada. Käesoleva projektiga soovime selle tühimiku täita luues analüüsivahendi, mida poliitikakujundajad saavad kasutada sotsiaaldemograafiliste stsenaariumanalüüside läbiviimiseks ning millel on potentsiaal kasvada ja edasi areneda.

Projekti veab Tartu Ülikool, tehes seda koostöös Eesti Rakendusuuringu Keskusega CentAR. Tartu Ülikoolil on pikk kogemus rahvastiku regionaalsete prognooside koostamises ning tervisenäitajate analüüsimises. CentAR panustab projekti lisaks oma sotsiaalmajanduslike analüüside läbiviimise kompetentsidele kogemusega arendada mikrosimulatsioonimudeleid LIAM2 keskkonnas.

1.2. Generatiivse tehisintellekti kasutamine taotluse koostamisel

Kas te olete kasutanud generatiivse tehisintellekti abi (nt ChatGPT, MS Copilot, Gemini, DALL-E, Claude vms) taotluse või selle osade kirjutamiseks või ettevalmistamiseks?

Jah

1.2.1. Selgitus tehisintellekti kasutamise kohta

Uurimismeeskond kasutab oma töös peamiselt ühte generatiivsel tehisintellektil põhinevat abivahendit - ChatGPTd. Taotluse kirjutamisel kasutati seda peamiselt:

- Ideede valideerimiseks
- Taotluse teksti struktuuri parandamiseks
- Õigekirja kontrolliks
- Analüüsikoodide valideerimiseks.

2. Uuringukirjeldus

2.1. Pakkuja arusaam sisulisest uuringuprobleemist ja sellest lähtuvalt püstitatud uurimisülesannetest

Tellijal on uurimisprobleemi määratlenud järgmiselt: **Eestis puudub pidevas kasutuses ja arenduses olev teaduspõhine mikrosimulatsioonimudel, mis võimaldaks simuleerida tuleviku arenguid ja oleks piisavalt laiapõhjaline, et toetada erinevate eluvaldkondade poliitikanalüüse.**

Uurimisprobleem esimene osa adresseerib olemasolevate mikrosimulatsioonimudelite piiratust. Eestis on mõned sotsiaalvaldkonnas poliitikakujundamisel kasutatavad mikrosimulatsioonimudelid – eeskätt EUROMOD ja pensioni mikrosimulatsioonimudel – kuid nende skoop ei vasta uurimisprobleemi kirjeldusele. EUROMOD on muutujate poolest küllaltki laia haardega, kuid staatilise arhitektuuriga. Selle abil on võimalik simuleerida poliitikamuudatuste (nt muutused maksusüsteemis või toetustes) mõju elanikkonnale enne-pärast raamistikus. Dünaamilisi simulatsioone ei ole võimalik staatilise mudeliga läbi viia. Pensionimudel on ülesehituselt dünaamiline ja seda on võimalik kasutada lisaks pensionisüsteemile ka rahvastiku ja tööturu simuleerimiseks, kuid, nagu Tellija poolses probleemikirjelduses välja toodi, on muutujate ring, mida rahvastiku simulatsioonides (ning nendel põhinevates tööturu ja pensionisimulatsioonides) arvesse võetakse täna liiga kitsas ning ei hõlma mh ei regionaalset paiknemist, leibkondlikku kuulumist ega tervisenäitajaid.

Eestis on hea kvaliteediga registriandmestikke, mida saaks kasutada rahvastiku mikrosimulatsioonimudelile täiendavate dimensioonide lisamiseks (nt saaks Rahvastikuregistrit

või Statistikaamet residentsusindeksi põhised infot kasutada simulatsioonidele regionaalse dimensiooni lisamiseks, Töötukassa, Tervisekassa ja surmaregistrites olevat infot saaks kasutada inimeste töövõime ja/või tervises seisundi integreerimiseks mudelisse ning mingil määral oleks võimalik kasutada elukohainfot ka leibkondade konstrueerimiseks). Seega on uurimisprobleem meie hinnangul õigesti püstitatud - mikrosimulatsioonimudelite maastikul on tühimik, mis vajaks täitmist ning eksisteerivad ka andmestikud, mille põhjal on võimalik (eeldusel, et AKI ja eetikakomitee annavad selleks loa) puuduvaid mõõtmeid mikrosimulatsioonimudelitesse lisada.

Uurimisprobleemi teine osa puudutab erinevate teadusprojektide raames loodavate andmestike ja töövahendite jätkusuutlikkust. Mudelid vajavad pidevat uuendamist ja arendamist, kuid pärast projekti lõppemist on seda üldjuhul keeruline korraldada – seda nii seetõttu institutsionaalne raamistik ei ole ehitatud selliselt, et toetada mudeli kestmist ka pärast projekti lõppu (nt andmekaitse küsimused), aga ka selle pärast, et mudeli arendamise ja kasutamise kompetents on koondunud liiga kitsa ekspertide grupi kätte. Kindlasti pole vähetähtis ka see, et mudelite uuendamiseks ja arendamiseks pole sageli planeeritud jätkuressurssi. Selleks, et mudelid püsiks „elus“ ja oleks heaks sisendiks poliitikakujundamisse, peaks nende kasutajate ring olema võimalikult lai.

Eelpool kirjeldatud uurimisprobleemi lahendamiseks püstitatakse järgmised uurimisülesanded:

1. Tellija vajaduste ja ootuste täpsustamine
2. Mikrosimulatsioonimudeli arhitektuuri loomine
3. Mudeli baasandmestiku ja käitumuslike seoste loomine
4. Mudeli kodeerimine
5. Mudeli testimine ja täiendamine
6. Mudeli jätkusuutlikkust toetava raamistiku loomine

Tellija vajaduste ja ootuste täpsustamine

Enne mudeli ehitamise juurde alustamist tuleb täpsustada, millised on Tellija ootused mudeli funktsionaalsusele. Kuna ressursside piiratuse tingimustes tuleb teha kompromisse, siis arutatakse selles faasis Tellijaga läbi, milline on erinevate funktsionaalsuse komponentide prioriteetsus.

Pakkumiskutses nähtub, et mudel peab olema ühilduv Pensioni prognoosimudeliga, mis määrab mõningad infotehnoloogilised valikud (mikrosimulatsioon peaks toimuma tarkvaras LIAM2, andmete ettevalmistus oleks mõistlik teostada statistikatarkvaras R). Pensionimudeli rahvastiku moodulile lisanduvate täiendavate mõõtmete osas on pakkumiskutses kirjas, et mikrosimulatsioon võiks hõlmata võimalust simuleerida Eesti sisest regionaalset paiknemist, leibkonda ja tervises seisundit). Sellele vaatamata on mitmeid teemasid, mis tuleks tellijaga enne mudeli arendamise juurde asumist läbi arutada:

- a) Kuidas plaanib tellija mudelit kasutama hakata?
- b) Millised peaks olema mudeli simuleeritavad tunnused?
- c) Millise detailsusega peaks mudeli võimaldama simuleerida pakkumiskutses välja toodud regionaalset paiknemist ja tervises seisundit?
- d) Mille alusel oleks otstarbekas defineerida leibkondlik kuuluvus?

- e) Milliseid sisendparameetreid peaks kasutajal olema võimalik muuta?
- f) Millise rahvastiku prognoosiga peaks mudeli ühilduma (nt Eurostat, Statistikaamet)?
- g) Milliste andmestike põhjal tuleks mudeli baasandmestik üles ehitada?
- h) Kuidas tagada võimalikult laia kasutajaskonna juurdepääs mudelile?
- i) Kuidas tagada mudeli baasandmestiku ja käitumuslike seoste uuendamine võimalikult väikese ressursikuluga?

Mikrosimulatsioonimudeli arhitektuuri loomine

Kui tellija vajadused on selged, siis tuleb paika panna mudeli arhitektuur – määratleda, mida on võimalik mudeliga simuleerida ning kuidas mudel töötab. Muuhulgas tuleb selles etapis selgeks saada, milliseid andmeid mudel analüüside teostamiseks vajab, milliseid käitumuslike võrrandeid ja millisel kujul hinnatuna on mudeli toimimiseks vaja ning millised joondamistabelid tuleb koostada. Mudeli esialgne arhitektuur jääb lõplikuga võrreldes üldisemaks, töö käigus selguvad paratamatult need osa plaanist, mis hästi ei tööta ja vajavad täiendamist.

Mudeli baasandmestiku ja käitumuslike seoste ettevalmistus

Mudeli baasandmestiku, käitumuslike seoste ja joondamisreeglite ettevalmistamisel on oluline, et need oleks ühest küljest teoreetiliselt põhjendatud ning teisalt et need ühtiksid dünaamilise mikrosimulatsiooni raamistikuga. Teoreetiline põhjendus tähendab eeskätt seda, et peab olema põhjus eeldada erinevusi huvi pakkuvate gruppide sotsiaaldemograafilistes näitajates ning teooriast on abi ka nendele erinevustele funktsionaalse kuju andmisel (nt suremuse vanuskordajate käitumine üle vanuse). Koosõla dünaamilise mikrosimulatsiooni raamistiku vajadustega tähendab, et viis, kuidas gruppide vahelisi eripärasid modelleeritakse, peab tagama simuleeritavate indiviidide usutavad jaotuse nii vaatlusalusel aastal kui üle simulatsiooniperioodi.

Kõige paremini on see ilmestatav näitega tööturul, kus, nii nagu elus, peab simulatsioonis esinema inimese sissetuleku tasemes üle aastate mõningane juhuslik kõikumine. Samas ei saa sissetulekud olla täiesti juhuslikud, sest üldjuhul toimub kõrgema sissetulekuga inimeste sissetulekute kõikumine kõrgema ning madalama sissetulekuga inimestel madalama taseme juures. Lisaks ei tohi simulatsioon põlistada kõrgema sissetulekuga inimesi kõrgele ja madalamatele sissetulekuga inimese madalamale tasemetele, sest ka see poleks reaalsusega koosõlas. Selles arendusfaasis pole töö põhifookus tööturul, kuid sarnased kaalutlused kehtivad ka tervisenäitajate ja leibkonna kuuluvuse modelleerimisel.

Eeltoodu teeb mikrosimulatsiooni jaoks käitumuslike seoste ja joondamisreeglite hindamise keeruliseks ning neid eripärasid tuleb kindlasti arvestada.

Mudeli kodeerimine

Mudeli kodeerimine ning baasandmestiku, käitumuslike seoste ja joondamisreeglite ettevalmistus toimuvad paralleelselt. Mudeli üldisema kirjeldamise järel valmistatakse ette baasandmestik (kõigi Eesti inimeste kõik simulatsiooni läbiviimiseks vajalikud tunnused simulatsiooni algusaastal) ning hinnatakse ära käitumuslikud seosed ja joondamisreeglid (nt millise tõenäosusega liiguvad inimesed vaatlusalusel aastal ühest Eesti maakonnast teise, millise tõenäosusega muutub nende tervises seisund või leibkondlik kuuluvus). Kui baasandmestik ja käitumuslikud seosed on valmis, integreeritakse need mudelisse.

Mudeli testimine ja täiendamine

Käesoleva teadusprojekti tulemuseks peaks olema tellija vajadustele vastav töövahend. Selleks tuleb tellija kaasata mudeli väljatöötamise ja testimise. Seda tuleks teha nii varajases faasis kui võimalik.

Mudeli jätkusuutlikkust toetava raamistiku loomine

Mudeli jätkusuutlikkuse tagamise tegevused võib üldjoontes jagada kaheks. Esiteks aitab mudeli kasutamisele kindlasti kaasa, kui mudel on hästi dokumenteeritud ning varustatud kasutusjuhiseiga.

Lisaks sellele on oluline, et mudel paiknes kohas, kus see on laiemale teadlaskonnale kättesaadav ning teadlaskond on teadlik mudeli olemasolust ja selle kasutusvõimalustest.

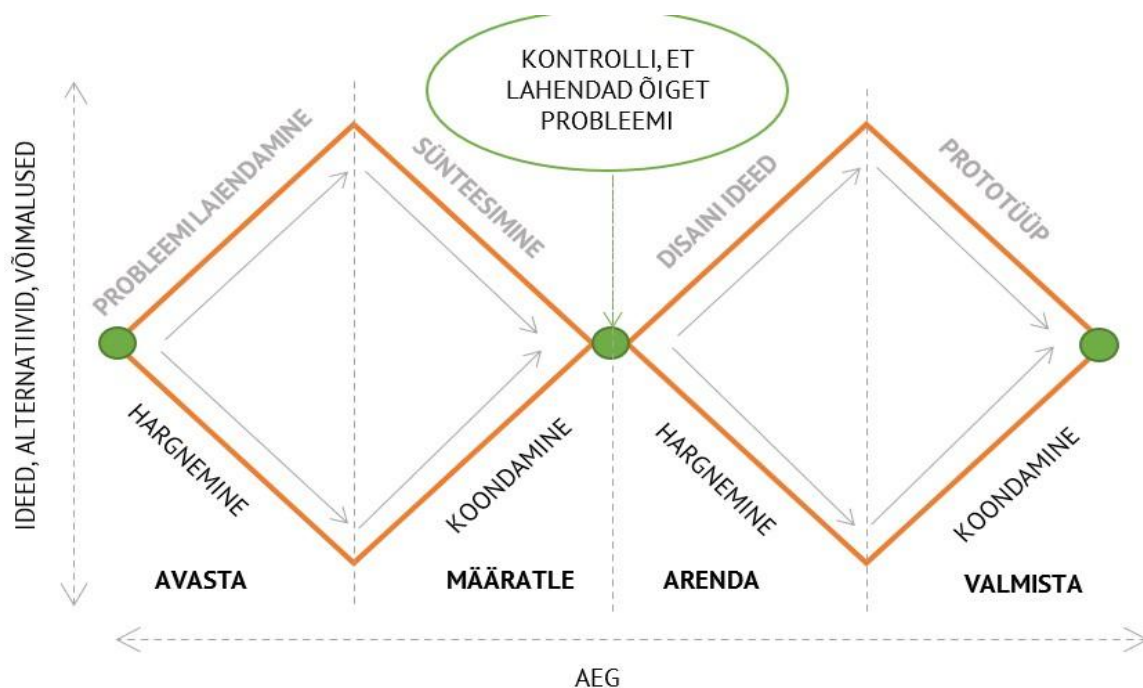
2.2. Pakutava uurimismetoodika kirjeldus ja põhjendus

2.2.1. Tellija vajaduste ja ootuste täpsustamine

Tellija vajaduste kaardistamisel kasutatakse kahekordse rombi (*double diamond*) metoodikat:

- Probleemi analüüs – kaardistatakse lahendamist vajavad probleemid.
- Süntees – täpsustatakse, millist probleemi lahendatakse.
- Ideede loomine – genereeritakse erinevaid lahendusvõimalusi.
- Lahenduse arendamine – valitakse ja viimistletakse sobivaim lahendus

Joonis 1. Teenusedisaini kahe rombi protsess



Sama lähenemist kasutas CentAR ka pensionimudeli pilootprojekti. Selle abil sõnastatakse põhiprobleemid ja tehakse metoodilised otsused. Arendustöö käigus toimuvad regulaarsed Teamsi-kohtumised tellijaga (eeldatavalt kord kuus), et anda ülevaade töö edenemisest ja lahendada jooksvaid küsimusi. Lisaks täpsustatakse tellijaga simulatsioonimudeli ühilduvust käimasoleva liikuvusmudeli koostamisega TÜ Mobiilsusuuringute Labori juhtimisel.

2.2.2. Rahvastiku dünaamilise mikrosimulatsioonimudeli loomine

Sissejuhatus

Rahvastikuprognosid jagunevad neljaks: trendi ekstrapoleerimine, kohort-komponendi meetod, struktuurimudelid ja mikrosimulatsioon (Smith, Tayman ja Swanson 2013). Mikrosimulatsiooni kasutatakse, kui soovitakse saada infot jaotuste kohta (nt kuidas jaotuvad palgad ühiskonnas). Dünaamiline mikrosimulatsioon võimaldab jälgida iga indiviidi elukäiku, hõlmates rohkem tunnuseid ja keerukamaid käitumismustreid kui kohordimudelid (Jia, Leknes ja Løkken 2023). Näiteid: Norra MOSART (Andreassen *et al* 2020), Kanada Demosim (Morency *et al* 2025), Austraalia APPSIM (Keegan 2007), Saksamaa MikroSim (Münnich *et al* 2021).

Pensionimudeliga ühilduv rahvastikumudel peaks olema dünaamiline mikrosimulatsioonimudel, mis (Järve *et al.*, 2023):

- kasutab dünaamilist vananemist,
- ristlõikelist simulatsiooniprotsessi,
- opereerib diskreetses ajas,
- tugineb registriandmetele,
- on suletud mudel.

Detailne metoodika selgub pärast tellijaga arutelusid, kuid olulised mõõtmed, mis pensionimudeli rahvastiku osas on puudu ja vajavad sisustamist on regiooni-, leibkonna- ja tervise mõõde.

Regionaalne mõõde

Regionaalsuse sisse toomisel erituvad kaks peamist teemade ringi:

- a) piirkondlikud erinevused sündimuses ja suremuses ning
- b) ränne (riigisisene ja rahvusvaheline).

Enamik dünaamilisi rahvastiku mikrosimulatsioone on loomult ülalt-alla mudelid, kuna nad kasutavad sisendina riigi tasandi andmetest tuletatud näitajad (sündimust, suremust, rändetöenäosusi) (Bacon *et al.* 2023). Samas – piirkondlikke eripäradeta jääb regionaalne käsitlus infovaeseks (Weymeirsch, Ernst, and Münnich 2024). Mõistlik on kasutada hübriidlahendusi. Kooskõla riigi tasandi prognoosidega tagatakse üldjuhul joondamistehnikatega (*alignment*) (vt nt Li ja O'Donoghue (2013)). Meie poolt eelistatud lähenemine on järgmine:

- a) Suremuse hindamisel kasutatakse riigi tasandi andmeid, pannes suremuse sõltuvusse tunnustest nagu sugu, vanus, haridus, tervisetegurid. Regionaalsed erinevused tulenevad erinevustes elanikkonna koosseisust.

- b) Kui selline lähenemine ei kirjelda piisavalt piirkondlikke erisusi, lisatakse regiooni-spetsiifilised efektid.
- c) Agregaatsed suremuse tõenäosused joondatakse riigi rahvastikuproгноosi järgi.
- d) Sündimus leitakse, võttes arvesse KOV tasandi efekte.

Kui pikaajaliselt on regionaalse demograafia olulisemateks mõjutajateks sündimus ja suremus, siis lühiajaliselt on selleks **ränne** (Vanella, Hellwagner, and Deschermeier 2023). Samuti võib KOV tasandil olla rände mõju oluliselt suurem kui sündimisel ja suremusel.

Rahvusvahelise- ja siserände modelleerimisel on sarnasusi. Regioonist lahkumist (olgu siis teise KOVi või välismaale) saab ajalooliste andmete põhjal modelleerida, hinnates lahkumise seost isiku vanuse, soo, hariduse ja teiste sotsiaaldemograafiliste näitajatega.

Lisaks lahkumise tõenäosusele on vaja määrata koht, kuhu rännatakse. Selleks kasutatakse rändemaatriksit – iga KOVi kohta saab ajaloole tuginedes öelda, milliste KOVide vahel inimesed peamiselt kolivad. Sama maatriks võib sisaldada ka tõenäosust kolida välismaale. Rände modelleerimisel võib olla mõistlik grupeerida KOVid tüübitunnuse järgi. Näiteks, mitte-tõmbekeskuse tüüpi KOVist liigutakse tõenäolisemalt tõmbekeskuse tüüpi KOVi.

Sise- ja välisrände käsitlemisel on ka eripärasid. Siseränne peab olema KOViti tasakaalus. Väljarändajad eemaldatakse üldjuhul rahvastikust. Lisaks ei tea me rahvusvahelise sisserändaja kohta kõiki isikutunnuseid (siserände isikutunnuste komplekt on rikkalikum, tunnustel on vähem puuduvaid väärtusi). Selle probleemi lahendamiseks kasutatakse „kloonimist“ – sisserändajatele antakse mõne Eesti inimesega sarnased tunnused (Dekkers 2015).

Leibkonna mõõde

Leibkond on oluline sotsiaalpoliitika analüüsimiseks, näiteks sotsiaaltoetuste arvutamisel. Mikrosimulatsioon võimaldab arvesse võtta infot, mida makromudelitesse ei saa lihtsalt integreerida. Näiteks, kuidas pärast vanemate surma jagunevad ülalpeetavad lapsed (Spielauer, Horvath, and Fink 2020).

Leibkonnad tulevad simulatsiooni sisse kahel moel:

- a) Leibkonna määratlemine
- b) Rände modelleerimine.

Leibkondi on võimalik modelleerida erinevalt. **Eksplitsiitse** modelleerimise korral simuleeritakse otseselt sündmusi nagu partnerluse moodustumine (abiellumine või kooselu alustamine), laste sünd jne. Mudeli andmestikus on isikutevahelised seosed. **Implitsiitse** lähenemise korral võidakse omistata igale isikule leibkonna staatus (nt “*elab üksinda*”, “*elab lapse/vanemaga*”) vastavalt tõenäosustele, ilma isikute vahele seoseid loomata.

Teiseks eristuseks on suletus vs avatus. **Suletud mudel** tähendab, et partnerlussuhted (kooselud) luuakse andmestikus olevate isikute vahel. See võimaldab mh järgida sugulussidemeid üle aja (Zagheni 2015). **Avatud mudeli korral** ei pea partner tulema andmestikust, ta luuakse vastavalt vajadusele. Lähenemise plussiks on tehniline lihtsus, miinuseks ebarealistlikum simulatsioon ja halvem kooskõla makrotasandi näitajatega (Li and O’Donoghue 2013).

Eesti mudelis:

- a) LIAM2s on partnerlussuhete tehniline modelleerimine lihtne, mistõttu tasuks kasutada suletud lähenemist.
- b) Implitsiitse ja eksplitsiitse lähenemise valik sõltub andmetest – kui leibkondade kohta on info olemas, siis võiks kasutada eksplitsiitset lähenemist.

Leibkonnad on olulised ka rände modelleerimisel – rändajad liiguvad tihti leibkonnadena (Dekkers 2015). Pensionimudeli rände osa on üles ehitatud indiviidi põhisel. Leibkonna sisse toomisel tuleks see integreerida ka rände modelleerimisse.

Tervise mõõde

Tervise sissetoomine võimaldab analüüsida tervise, tervishoiuteenuste kasutamise, haigestumise ja suremuse ning sotsiaalmajanduslike tulemuste mitmepoolset seost. Samuti on võimalik simuleerida (tervise)poliitika mõju nii tervisele kui sotsiaalmajanduslikele näitajatele. Üks olulisemaid otsuseid, mis tuleb langetada on: milleks plaanitakse mudelit kasutada (Spielauer 2007).

Mitmeotstarbelistes mudelites on tervisega seotud aspektid sageli kaasatud üsna algelisel kujul. Mudelid arvestavad tervist tavaliselt puude kaudu, mida kasutatakse peamiselt sotsiaalkindlustushüvitiste, institutsionaliseerimise vajaduse hindamiseks või tervishoiuteenuste vajaduse määramiseks ning mõnel juhul ka tervishoiukulude ja -kasutuse prognoosimiseks. (Zucchelli, Jones & Rice 2012). Info tervise kohta võimaldab täpsemalt prognoosida nt variatsiooni suremuses, tööturul lahkumises ja töötasus (nt Marois & Aktas, 2021) ning meie andmetes ka soovi korral sotsiaalhüvitiste kasutamist kui ka äri- ja kinnisvara muutusi indiviidi tasandil.

Mudelites, kus tervisepoliitika sekkumistel on primaarne roll, on vaja sisse tuua terviseseisundid detailsemalt, näiteks peamiste haigusrühmade esinemine (südame-veresoonkonna haigused, diabeet, vähid jm) ja lisatakse peamised riskitegurid ja nende muutumine (nt suitsetamine, alkoholi tarbimine, kehamassiindeks, vererõhk jm). Samuti sisaldavad need mudelid ka tervishoiuteenuste kasutamist. Taolistes mudelites on tervisepoliitika sekkumised suunatud riskitegurite mõjutamiseks ja tervishoiuteenuste vajaduse katmiseks.

Koostöös tellijaga täpsustakse, milliseid simulatsioone peab mudel võimaldama.

Üleminekuid terviseseisundite vahel modelleeritakse sarnaselt teiste tunnustega Markovi mudeli või diskreetse aja sündmuse mudeli abil, arvestades demograafilisi näitajaid (sugu, vanus, maakond), sotsiaalmajanduslikku staatust (palk, haridus) ja varasemaid terviseseisundeid (nt enesehinnang). Populatsiooni terviseseisundite algne jaotus leitakse kas rahvaloenduse andmetest või Eesti Tervisekassa ja Tervise infosüsteemi andmetest. Kui mikroandmete tasemel ei ole võimalik andmeid ühendada, siis simuleeritakse tervise- ja sotsiaalmajanduslike tunnuste ühisjaotused eraldiseisvate registriandmete või küsitlusuuringute põhjal. Aastane ümberkaalumine joondab simuleeritud levimuse, suremuse ja riskitegurite jaotused välise sihtväärtustega.

2.2.3. Mudeli baasandmestiku ja käitumuslike seoste loomine

Mudeli baasandmestiku koostamisel ning käitumuslike seoste ja joondamisreeglite arvutamisel kasutatakse erinevaid statistilisi ja ökonomeetrisi meetodeid, mille täpne valik sõltub mudeli

arendamisel tehtavatest otsustest. Kuna mudeli sisu on veel täpsustamisel, siis ei ole võimalik andmetöötluse metoodikat detailsemalt avada.

Andmeanalüüsiks on plaanis kasutada statistikatarkvara R.

2.2.4. Mudeli testimine ja täiendamine

Mudeli esmaversiooni valmimisel võimaldatakse tellijal mudelit testida (seda vajadust tuleb silmas pidada ka AKI ja eetikakomitee lubade taotlemisel). Projekti eelarves ja ajakavas on reserveeritud aeg nii testimiseks kui muudatuste tegemiseks.

2.2.5. Mudeli jätkusuutlikkust toetava raamistiku loomine

Mudeli jätkusuutlikkuse tagamiseks tehakse järgmisi tegevusi:

- a) Mudel plaanitakse majutada Statistikaametisse
- b) Mudeli kood ja andmete ettevalmistus dokumenteeritakse
- c) Mudeli kood on varundatud git repositooriumis
- d) Koostatakse mudeli kasutusjuhend
- e) Luuakse mudeli videotutvustus
- f) Mudelit populariseeritakse teadusartikli, sotsiaalmeedia ja avalike seminari kaudu
- g) Kaalutakse alusandmete värskendamise koostöökokkuleppe võimalust.

2.3. Pakkuja nägemus andmevajadusest ning andmestike sidumisest

Rahvastiku mikrosimulatsioonimudeli arendamiseks on vaja kahte erinevat tüüpi andmestikke:

- a) Simulatsiooni baasandmestik – andmestik, mis sisaldab simulatsiooni baasaastal (nt 2025. aasta) kõigi Eesti inimeste kõiki simulatsioonimudeli simuleeritavaid tunnuseid (sugu, vanus haridus jne).
- b) Simulatsioonimudeli arendamise alusandmestikud – need on andmestikud, millest tuletatakse grupipõhised parameetrid (nt seisundi leviku ulatus; ühest seisundist teisse liikumise tõenäosus) millele tuginedes hinnatakse mudelis kasutatavaid indiviidide käitumist kirjeldavad võrrandid. Ka mudeli baasandmestik luuakse alusandmestike põhjal. Vastavad alusandmestikud on muuhulgas järgnevad:
 - a. Teaduse teekaardi objekt Infotehnoloogiline mobiilsusobservatoorium (IMO ., <https://imo.ut.ee/>) hõlmab peamisi demograafilisi ja sotsiaalseid isikutunnuseid. IMO andmebaas asub Statistikaameti teadlaste turvalisel töökohal. Nendele andmetele palutakse projekti täitjatel ligipääsu Statistikaameti konfidentsiaalsusnõukogult. Nendele andmete on vajadusel juurde lingitavad ka teised Statistikaameti andmestikes olevad andmed.
 - b. BIG-HEARTi andmebaasis (vt täpsemalt Lõo *et al* (2025)) on kõik Eesti elanikud vanuses 36+ aastal 2012 (N=770,323) kelle terviseseisundit jälgitakse 12 aastat. Uuringu on heaks kiitnud TÜ Inimuuringute Eetikakomitee (otsus 390/M-19; 20.01.2025). Peale käesoleva uuringu alustamist taotleme loa laiendamist uueks otstarbeks mida plaanime saada 2 kuu jooksul. Pseudoisikukood andmete sidumiseks (ka ema, isa ja laste isikukoodid peavad selle põhised olema, võimaldama inimesi omavahel linkida). Oluline on siiski mainida, et BIG-HEARTi

andmestik asub Tartu Ülikooli teadusarvutuskeskuse turvalises serveris SAPU ning neid andmeid ei saa siduda Statistikaameti andmetega. Küllaga on võimalik kasutada seda, väga rikkalikku, andmebaasi selleks, et hinnata seoseid tervise seisundi ja muude sotsiaaldemograafiliste näitajate vahel.

Andmekoosseis sisaldab muuhulgas:

- i. Rahvastikuregistrist sugu
- ii. Sünniaasta (sünnikuu on puudu)
- iii. Surma kuupäev
- iv. Kodakondsus ja kõrgeim omandatud haridustase ISCED2011 klassifikaatoriga; mõlemad on ühekordse 2023 seisuga väljavõtted rahvastikuregistrist). Sarnaselt on olemas rahvus ja abielustaatus.
- v. Kohalik omavalitsus, kus inimene elab.
- vi. (Puudu on andmed Vanemate ja laste kohta; sünniriik; riigist lahkumine/sisenemine)
- vii. Töötüregistrist ja Sotsiaalkaitse infosüsteemist sisuliselt kõik kanded perioodil 01.01.2008–31.12.2015, sh töötutoetus; töötukindlustushüvitis; kutsehaigusest/tööõnnetusest tingitud tervisekahju hüvitis; puuetega inimeste sotsiaaltoetus; töövõimetuspension; kuriteoohvri hüvitis; represseeritu toetus; toetus Tšornobõli katastroofi tagajärgede likvideerijale; üksi elava pensionäri toetus.
- viii. Kinnistusraamatust ja Äriregistrist ülevaade millist vara uuritav omas 01.2012, 01.2014 ja 01.2016 kuupäevadel.
- ix. Tervisekassa andmekogust kõik raviarved (sh perearstid ja haiglad), välja kirjutatud ja/või ostetud ravimid vahemikus 01.2010–12.2023. Antud andmed on omakorda viidud üle rahvusvaheliselt tunnustatud OMOP-CDM standardsele andmekujule (Oja *et al* 2023)), mis hõlbustab kodeerimist ja analüüsimist (nt rinnavähk jah/ei).

Andmestike kasutamiseks on vaja eetikakomitee ja võimalik et ka AKI luba (kuna viimasel ajal on erinevate lubade vajaduste praktika olnud hektiline, siis täpsustub lubade vajadus projekti käigus). Need taotletakse pärast seda, kui tellijaga on kokku lepitud mudeli funktsionaalsus. Tõenäoliselt on AKI nõusolek ja eetikakomitee luba vajalikud teiste andmestike kasutamiseks.

CentAR on taotluse esitamise ajal lõpetamas Pensionimudeli arendamist. Seal võeti suund, et mudeli baasandmestiku ja seoste uuendamiseks peaks simulatsioonimudeli arendamise alusandmestik paiknema Statistikaametis ning mudeli arendamise käigus koostatakse andmetöötlusskriptid, mida on võimalik kasutada baasandmestiku uuendamiseks nii, et analüütiku sekkumine sellesse protsess on minimaalne.

Meie hinnangul on ka selle mudeli arendamisel tark kasutada sama taktikat ning hoida simulatsioonimudeli arendamise alusandmestikku Statistikaametis. Töö andmestikuga peaks toimuma Statistikaameti turvalises keskkonnas (üle interneti) nii, nagu see on lahendatud IMO andmetega.

2.4. Pakkaja nägemus uuringuprotsessi toimimisest

Üldine korraldus

Kogu projekti vältel korraldatakse tööseminare, kus osalevad tellija ja teostaja esindajad. Need toimuvad igakuiselt (vajadusel võidakse kokku leppida erandid). Tööseminaridel annab teostaja ülevaate tehtud töödest, tekkinud väljakutsetest ning nende lahendustest. Tellija annab tehtule tagasisidet ning räägib kaasa metoodiliste valikukohtade või ilmnunud väljakutsete lahendamisel. Tööseminarid toimuvad vastavalt kokkuleppele tellijaga kas Teamsis või silmast silma. Tellijale esitatakse 2 korda aastas kirjalik vahekokkuvõtte tööde käigust.

Temaatiline korraldus

Uurimistöö on jagatud kuueks sisuliseks töopaketi, millele lisandub juhtimine:

1. Tellija vajaduste ja ootuste täpsustamine (TP1)
2. Mikrosimulatsioonimudeli arhitektuuri loomine (TP2)
3. Mudeli baasandmestiku ja käitumuslike seoste loomine (TP3)
4. Mudeli kodeerimine (TP4)
5. Mudeli testimine ja täiendamine (TP5)
6. Mudeli jätkusuutlikkust toetava raamistiku loomine (TP6)
7. Uuringu juhtimine (TP7)

Tellijaja vajaduste ja ootuste täpsustamine (TP1)

Kaks seminari, kus arutatakse tellijaga läbi nende vajadused. Muuhulgas täpsustatakse, milline peaks olema mudeli simulatsioonivõimekus (milliste parameetrite mõju peaks saama simuleerida) ja millist regionaalse jaotuse detailsust soovitakse mudelis näha. Tööpaketi tulemused vormistatakse lühiraportina.

Vastutus: Põhivastutus CentARil, TÜ rahvastiku ja tervise tiimid osalevad aruteludel ja toetavad oma ekspertiisiga.

Mudeli arhitektuuri koostamine (TP2)

Aluseks võetakse pensionimudel. Seda analüüsitakse koostatakse uus tellija vajadustega haakuv arhitektuur. Muu hulgas määratakse ära, milliseid sisendandmeid mudel analüüside teostamiseks vajab, milliseid käitumuslike võrrandeid ja millisel kujul hinnatuna on mudeli toimimiseks vaja ning millised joondamistabelid tuleb koostada.

Vastutus: Põhivastutus on CentARil, kus on sarnase mudeli ehitamise kogemus. Samas kasvab siit välja valdav osa edasisest tööst mis tähendab, et kogu meeskond peab olema hästi kursis sellega, mida ja miks mudeli sisendiks vaja on.

Mudeli baasandmestiku, käitumuslike seoste ja joondamistabelite loomine (TP3)

Selles etapis toimub:

- Mudeli alusandmestiku määratlemine (andmed, mida on vaja baasandmestiku, käitumuslike võrrandite ja joondamistabelite koostamiseks) (CentAR, TÜ)

- Andmete töötlemiseks loa taotlemine AKIlt ja eetikakomiteelt (BIG-HEART andmestiku puhul on vaja vaid loa muudatust) (CentAR)
- Mudeli baasandmestiku koostamine (pseudonümiseeritud andmed iga inimese kohta, mida mudeli simulatsioonide läbiviimiseks kasutatakse) (CentAR, TÜ)
- Käitumuslike võrrandite hindamine (CentAR, TÜ)
- Joondamistabelite koostamine (CentAR, TÜ)

Eeldatavalt tuleb sisulises lõikes lahendada selles etapis järgmised ülesanded:

- Sisustada hariduse muutuja ja hariduse omandamise protsess (mis ja kuidas mõjutab hariduse omandamist) (vastutab CentAR)
- Sisustada tervise muutujad ja terviseseisundi muutumise protsess (mis ja kuidas seda mõjutab) (vastutab TÜ tervis)
- Sisustada leibkonna tunnus ning määratleda ära leibkondade tekkimise, nendega liitumise ja sealt lahkumise protsessid (vastutab CentAR)
- Määratleda, kas ja kuidas mõjutavad haridus, tervis, leibkonna kompositsioon (võibolla veel mõni näitaja (nt sissetulek)), sündivust, suremust ja rännet (vastutab TÜ rahvastik).

Vastutus: Sisuteemade jaotuse eest vastutajad on toodud eelmises punktis. Lisaks eeltoodule toetab CentAR kogu protsessi LIAM2 alaste teadmistega (mis kujul peavad analüüsi tulemused olema, et sobida LIAM2 mudelisse).

Mudeli ehitamine (TP4)

Pannakse kokku mikrosimulatsioonimudel. Eelmise etapi tulemused tõlgitakse LIAM2 formaati, viiakse läbi esmased testsimulatsioonid ning tehakse selle põhjal korrekture. Suure tõenäosusega on vaja pärast testsimulatsiooni baasandmestikku, käitumuslike seoseid ja joondamistabeleid täiendada / ümber hinnata.

Vastutus: Mudeli kodeerimise vastutus on CentARil, TÜ aega on protsessi planeeritud arvestades vajadusega kohendada mudeli baasandmestikku, seoseid ja joondamistabeleid.

Mudeli testimine ja täiendamine (TP 5)

Mudeli tellijapoolne testimine ning selle põhjal mudeli täiendamine.

Vastutus: põhivastutus CentARil, võimalik andmete ja seoste ümbertegemine / hindamine CentARil ja TÜL.

Mudeli jätkusuutlikkust toetava raamistiku loomine (TP6)

Mudeli kasutamise jätkusuutlikkuse tagamiseks on plaanis teha järgmisi tegevusi:

- Mudel plaanitakse majutada Statistikaametisse
- Mudeli kood ja andmete ettevalmistus dokumenteeritakse
- Mudeli kood on varundatud git repositooriumis
- Koostatakse mudeli kasutusjuhend
- Luuakse mudeli videotutvustus
- Mudelit populariseeritakse teadusartikli, sotsiaalmeedia ja avalike seminari kaudu

- Kaalutakse alusandmete värskendamise koostöökokkuleppe võimalust

Vastutus: põhivastutus CentARil.

Uuringu juhtimine (TP7)

Projektijuhtimine, tellijaga tööseminaride korraldamine, vahekokkuvõtete ja lõppraportite vormistamine. Lähtub eeldusest, et kulu jääb alla 10% projekti eelarvest.

Vastutab: Projekti üldise juhtimise vastutus on TÜ-l, CentAR juhhib arendust.

2.5. Uuringu Gantti graafik

Paikneb eraldi failis (Lisa 2. Eelarve ja tööde jaotus.xlsx)

2.6. Riskide maandamise plaan

Peamised projekti realiseerimisega seotud riskid:

1. Pakkujate ja tellija nägemus uuringust ei lange kokku

- Esinemise tõenäosus: madal
- Olulisus: suur
- Vastutaja: projektijuht
- Selgitus: Võib juhtuda, et tellija ja pakkuja nägemused uuringust, sh uuringu eesmärkidest ja läbiviimisest, on erinevad.
- Maandamine: Aktiivne koostöö tellijaga nii uuringu algfaasis kui ka hilisemates etappides. Tellija vajaduste täpsustamiseks on kohe uuringu alguses ette nähtud eraldi etapp. Samuti on plaanis mudel välja arendada tihedas koostöös tellijaga (st tellija meeskonna liikmed osalevad arendusprotsessis).

2. Finantsrisk

- Esinemise tõenäosus: madal
- Olulisus: keskmine
- Vastutaja: projektijuht
- Selgitus: Võib tekkida olukord, kus konsortsiumil puuduvad piisavad rahalised vahendid projekti tegevuste läbiviimiseks, see muutub maksejõuetuks ning seetõttu jääb projekt teostamata.
- Maandamine: Riski tõenäosus on väga väike arvestades, et konsortsiumi üheks partneriks on Tartu Ülikool. Ka konsortsiumi teine partner (Centar) on pikaajalistelt kasutanud konservatiivset eelarvestamise poliitikat. Konsortsiumil on projektide sildfinantseerimiseks reservis piisaval hulgal rahalisi vahendeid.

3. Mõni projekti meeskonnaliikmetest lahkub pikemaks ajaks või jäädavalt töölt ja/või projektist

- Esinemise tõenäosus: keskmine
- Olulisus: madal
- Vastutaja: projektijuht

- Selgitus: Võib juhtuda, et mõni projekti meeskonnaliikmetest lahkub, sellega seoses kaob projektimeeskonnast kompetents ning sisuline võimekus analüüs edukalt lõpuni viia.
- Maandamine: Projektimeeskond on üles ehitatud nii, et meeskonnaliikme võimalikul haigestumisel/lahkumisel on olemas talle asendaja (tiimis on neljal inimesel, mikrosimulatsioonimudeli arendamise kogemus ja nad on võimelised teineteist asendama).

4. Projektijuht lahkub pikemaks ajaks või jäädavalt töölt ja/või projektist

- Esinemise tõenäosus: madal
- Olulisus: suur
- Vastutaja: projektijuht
- Selgitus: Projektijuhi haigestumisel või töölt lahkumisel võib kaduda ülevaade sellest, millised tegevused, kelle poolt ja mis ajaks peavad realiseeruma. Kaob terviku tunnetus projekti erinevatest osadest ja sellest, kuidas need omavahel kokku sobituvad.
- Maandamine: Projektijuhi võimalikul haigestumisel on projekti meeskonnaliikmete seas piisava projektijuhtimise kogemusega asendaja (Janno Järve), kes on jooksvalt kursis projekti seisuga ning saab teda vajadusel asendada. Meeskonnatööl põhinev töökultuur tagab ka selle, et erinevad meeskonnaliikmed teavad, kuidas nende töösad peaksid kokku sobima.

5. Tehnilised riskid: meeskonnaliikme arvuti vargus, purunemine, failide kogemata kustutamine, intervjuude salvestamine ei õnnestu

- Esinemise tõenäosus: madal
- Olulisus: keskmine
- Vastutajad: kõik meeskonna liikmed
- Selgitus: Meeskonnaliikme arvuti varguse, purunemise, failide kogemata kustutamise tulemusena läheb suur osa tööst kaduma, tähtajad venivad.
- Maandamine: Arvutite ning neis sisalduvate andmete kadumise riski maandame läbi automatiseeritud varundussüsteemi kasutamise (MS OneDrive). Andmete lekkimist väldime salasõnadega kaitstud arvutite ja nutiseadmete kasutamisega, andmeid töödeldakse Statistikaameti turvalises keskkonnas.

6. Mudeli arendamisega seotud riskid - sisendandmetega seotud võimalikud probleemid

- Esinemise tõenäosus: kõrge
- Olulisus: kõrge
- Vastutaja: projektijuht / tellija
- Selgitus: Üks peamisi riske on seotud mudelite arendamiseks vajalike andmete kättesaadavuse ja kvaliteediga. Kuna kasutatakse pseudonüümitud isikuandmeid, siis on vajalik AKI ja eetikakomitee luba, mille saamine ei ole ette teada olev sündmus. Lubade taotlemised võtavad aega, mis võib projekti venitada.
- Maandamine:

- Riski pole võimalik täielikult maandada, selle realiseerimine ning olulisus sõltub muuhulgas paljuski sellest, millist mudelit tellija lõpuks soovib. Peame siin tegema koos tellijaga tarku valikuid (seetõttu on ühe vastutajana välja toodud ka tellija).
- Mudeli arendajatel on aastatepikkune kogemus suhtlemisel eetikakomiteede ja AKIga. Nende nõusoleku saamine ei ole kindel sündmus, kuid selle tõenäosus on küllaltki.
- Oluliste andmebaaside (sh BIG-HEART) eetikaloa on juba olemas ning teiste vajalike andmete (IMO) kasutamise taotlemise kord on Statistikaametis välja töötatud. See maandab riski, et komitee ei kooskõlasta taotlust.

7. Mudeli arendamisega seotud riskid - metoodilised probleemid

- Esinemise tõenäosus: kõrge
- Olulisus: kõrge
- Vastutaja: projektijuht
- Selgitus: Mikrosimulatsioonimudeli arendamine eeldab keerukate ökonomeetriliste meetodite rakendamist mis ei pruugi alati anda soovitud tulemusi anda.
- Maandamine:
 - Meeskonna varasem kogemus pensionimudeliga kui ka tervisemudelitega – meil on hea ettekujutus sellest, milliseid probleeme võib tekkida ja kuidas neid lahendada.
 - Testida ja valideerida mudeleid hoolikalt, kasutades nii ajaloolisi andmeid kui ka stsenaariumianalüüse.
 - Säilitada projekti käigus paindlikkus, et vajadusel kohandada metoodikat vastavalt esile kerkivatele probleemidele.

2.7. Kommunikatsiooniplaan

Projekti kommunikatsioonitegevuste eesmärk on tagada teadustulemuste nähtavus nii akadeemilises kogukonnas kui ka ühiskonnas laiemalt. Mikrosimulatsioonimudeli arendus keskendub rahvastiku regionaalsete, tervise- ja leibkonnadimensioonide lisamisele Pensionimudeli rahvastiku moodulile ning on tihedalt seotud avaliku poliitika kujundamise ja teaduspõhise arutelu edendamiselega.

Kommunikatsiooniplaan arvestab, et projekti põhifookus on teaduslik arendus ning ressursid on piiratud. Seetõttu on rõhk valitud kanalite sihipärasel kasutamisel ja selgetel võtmesõnumitel.

Eesmärgid:

- a) Tutvustada projekti eesmärgi ja tulemusi poliitikakujundajatele, teadlastele ja praktikutele.
- b) Tõsta teadlikkust rahvastiku mikrosimulatsiooni kui poliitikainstrumendi olulisusest laiemale avalikkusele.
- c) Toetada projekti teaduslikku väljundit ja tagada vähemalt ühe teadusartikli avaldamine rahvusvahelises eelretsenseeritavas ajakirjas.

Sihtgrupid ja kanalid

- **Teadus- ja ekspertkogukond:** rahvusvahelised ja Eesti demograafia, majanduse ja rahvatervise teadlased; kanalid: teadusartiklid, konverentsiettekanded.
- **Poliitikakujundajad ja ametnikud:** Rahandusministeerium, Sotsiaalministeerium, Riigikantselei, Statistikaamet jt; kanalid: lühiraportid ja seminarid.
- **Laiem avalikkus:** huvilised ja arvamusiidrid; kanalid: ERR-i teadusportaal Novaator, Postimees, Delfi, Facebooki projektileht.

Tegevused ja ajakava

- Projekti algus (1–3. kuu):
 - Projekti eesmärkide tutvustus Facebooki kaudu.
 - ERR Novaatoris või muus portaalis lühike artikkel projekti algusest.
- Projekti vaheetapp (12.–15. kuu):
 - Populaarteaduslik artikkel Postimehes või Delfis, keskendudes regionaalsele või tervisedimensioonile.
 - Facebookis vahekokkuvõtte postitused (2–3).
 - Vaheseminarid (2 tk)
- Projekti lõpp (20.–24. kuu):
 - Teadusartikli avaldamine.
 - Lõpuseminar ekspertidele ja poliitikakujundajatele.
 - ERR Novaatoris või teises kanalis populaarteaduslik kokkuvõtte mudeli arenduse tulemustest.
 - Facebookis kokkuvõttepostitus.

Ressursikasutus

Tegevused planeeritakse ressursisäästlikult: populaarteaduslike artiklite tekstid koostatakse vahetulemite põhjal; Facebooki postitused (kokku ca 5) hoitakse kompaktsed. Meediakajastuse puhul otsitakse koostööd ERR-i ja suuremate portaalide toimetajatega, et vähendada lisatöökoormust.

Hindamine

Kommunikatsiooni edukust hinnatakse järgmiste näitajate alusel: teadusartikli aktsepteerimine, vähemalt 5 meediakajastust meediaväljaannetes, vähemalt 5 Facebooki postitust, lõppseminari toimumine.

Kuna selle punkti all on võimalik kasutada rohkem tähemärke, siis soovime juhtida tähelepanu sellele, et meetoodika kirjelduseks ette nähtud maht oli väga piiratud (10 000 tm), mistõttu mudeli

jätkusuutlikkuse tagamise osale oli võimalik pühendada vaid kompakne loetelu tegevustest. Püüame seda olukorda siin veidi parandada.

Meie nägemuses tagab mudeli jätkusuutlikkuse, kui teadlaskond ja ametnikud on mudelist huvitatud ja pääsevad sellele ligi. Kommunikatsioonil on siin oluline roll. Mõlema sihtrühma teadlikkuse tõstmiseks on plaanis seminarid (nii vahe- kui lõppseminarid), lisaks on teadusartikkel viidatav kõrge kvaliteediga allikas teadlaskonnale ning mudelit laiemalt populariseerivad tekstid (FB, artiklid ajakirjanduses) aitavad täiendavalt tõsta ametkonna huvi mudeli vastu. Kombineerides eeltoodut sellega, mudel hakkab paiknema Statistikaametis, kus ei teki vajadust mudeli baasandmestiku kustutamiseks projekti lõppemisel ning IMO andmestikel põhinev alusandmestik saab säilida ja uueneda vastavalt IMO reeglitele, tekivad head eeldused tagamaks mudelile pikk ja viljakas elu.

2.8. Uurimisteema panus RITA eesmärkide täitmisse

Projekti eesmärk on arendada Eesti jaoks rahvastiku mikrosimulatsioonimudel, mis võimaldab teaduspõhiselt simuleerida demograafilisi, regionaalseid, tervise- ja leibkonnaga seotud arenguid. Praegu puudub Eestis selline pidevalt kasutuses olev ja edasiarendatav tööriist, mis võimaldaks erinevaid poliitikavaldkondi integreeritult dünaamilises mikrosimulatsiooni raamistikus analüüsida ja prognoosida.

Projekti tulemused aitavad täita ühtekuuluvuspoliitika fondide rakenduskava prioriteedi „Nutikam Eesti“ erieesmärki „Teadus- ja innovatsioonivõime ning kõrgetasemeliste tehnoloogiate kasutuselevõtu arendamine ja suurendamine“, kuna suurendavad teadus- ja innovatsioonivõimekust ning loovad eeldused kõrgetasemeliste andmeanalüüsitehnoloogiate kasutuselevõtuks riigi ja kohalike omavalitsuste poliitikakujundamises. Spetsiifiliselt võimaldavad need väärindada meie rikkaliku registriandmestikku, mis täna on alakasutatud.

Projekt panustab otseselt teadmussiirde programmi 2024–2027 eesmärkide täitmisesse, mille kohaselt Eesti areng peab tuginema teaduspõhistele ja innovaatilistele lahendustele. Mikrosimulatsiooni kasutamine rahvastiku modelleerimiseks ja prognoosimiseks on kaasaegne lähenemine, mida võiks Eestis rohkem kasutada. Nii riigi kui kohalike omavalitsuste tasandil suureneks seeläbi võimekus hinnata reformide mõjusid tööjõu, tervise, hariduse ja sotsiaalkaitse valdkondades, paremat infot saadaks just selle kohta, kui võrdselt või ebavõrdselt on erinevad näitajad (nt keskmine palk, tervisenäitajad) Eestis jaotunud ning kuidas need jaotused aja jooksul muutuvad. Loodetavasti sünnivad sellest informeeritumad otsused.

Meede 21.1.1.2 seab eesmärgiks kasvatada ühiskonna ja majanduse vajadustele vastavat teadus- ja arendustegevuse võimekust. Projekti tulemuseks on uus teaduslik kompetents ja praktiline tööriist, mis panustab teadmiste põhisesse poliitika kujundamisse.

2.9. Mõju uuringuvaldkonna arengule Eestis

Eesti sotsiaalmajanduslike küsimuste analüüsimisel on dünaamilised mikrosimulatsioonimudelid küllaltki uus valdkond. Käesolev projekt võimaldab jätkata lähteülesandes viidatud Pensionimudeliga alguse saanud võimekuste kasvatamist Eestis.

Pensionimudeliga võrreldes on laienenud arendusse kaasatud organisatsioonide ring – seekord arendatakse mudeli Tartu Ülikooli ja CentARi koostöös. Lisaks laieneb ka valdkondlik haare, mida dünaamiline mikrosimulatsiooni mudel katab. Pensionimudel is käsitus leidnud, demograafia, tööturu ja pensioni teemad laienevad regionaalse analüüsi võimekuse, tervise ja leibkondade detailsema käsitus võrra.

Kui analüüsi käigus otsustatakse kasutada BIG-HEART andmesikku, siis tasub mainida, et tegemist on ühe maailma suurim ja parim andmebaas, mis ühildab endas nii tervise- kui sotsiaalvaldkonna andmed indiviidi tasandil. Selle kasutamine panustaks samuti märkimisväärselt valdkonna arengusse.

Lõpetuseks olgu mainitud, et konsortsium ei alatähtsusta järelkasvu. Projekti käigus kaalutakse võimalusi doktorandide ja magistrandide kaasamiseks. Taotluse kirjutamise faasis on siiski keeruline öelda, millises mahus seda lõpuks teha õnnestub.

3. Uurimismeeskond

3.1. Olen teadlik, et vajadusel tehakse uuringu meeskonnale taustakontroll

3.2. Projektijuht ja põhitäitjad

Konsortsiumi juht: Tiit Tammaru, Ph.D

Projekti juht: Janno Järve, Ph.D.

Põhitäitjad Tartu Ülikoolist:

- a) Andres Vörk, MA
- b) Taavi Tillmann, MBChB, PhD
- c) Tiit Tammaru, PhD.
- d) Kadi Kalm, PhD.
- e) Kaidi Nõmmela, PhD.

Põhitäitjad CentARist:

- a) Sten Anspal, Ph.D.
- b) Marko Sõmer, M.A.
- c) Laura Kivi, Ph.D.
- d) Janno Järve, Ph.D.
- e) Kati Kadarik-Trei, Ph.D.

3.3. Täitjad sh tudengid

Täitjad Tartu Ülikoolist:

- a) Kirils Gončarovs, MA.

3.4. Täitjate CV-d

Täitjate CVd on kättesaadavad ETISest.

3.5. Abitööjõud

Ei plaanita kasutada

3.6. Uurimisrühma pädevused ja uurimisrühma liikmete rollide kirjeldus

Lisatud eraldi failina (Lisa 1. Uurimisrühma liikmete pädevused ja rollid)

3.7. Allhankijate ülevaade

Projekti sisutegevustesse ei ole plaanis allhankijaid kaasata. Teatud mõttes võib allhankijaks lugeda Statistikaameti, kes pakub andmetele juurde pääsemiseks turvalise keskkonna teenust, kuid nende panus on tugiteenuse iseloomuga ja eeldatavalt ei vaja põhjalikumalt lahkamist.

3.8. Konsortsiumi kirjeldus

Käesoleva projekti elluviimiseks on vajalik küllaltki laia haardega meeskond, mille liikmed oleksid kogenud demograafiliste protsesside, tervise ja sotsiaalvaldkonna teemade analüüsimisel, valdaksid kõrgtasemel andmetöötlusmeetodeid ja suudaks kokku panna toimiva dünaamilise mikrosimulatsioonimudeli. Leiame, et meie meeskond on võimeline neid ülesandeid edukalt täitma.

Konsortsium koosneb Tartu Ülikooli ja CentARi ekspertidest, kellel on üksteist täiendavad teadmised ja kogemused. Meeskonna koosseis ja olemasolev taristu loovad tugeva aluse projekti edukaks elluviimiseks.

Tartu Ülikool toob projekti rahvastikuprotsesside modelleerimise kompetentsi. Ülikooli poolset tiimil on pikaajaline kogemus rahvastikuprognoside koostamisel nii riigi- kui regionaalsel tasandil. Lisaks on Tartu Ülikooli poolsetel meeskonnaliikmetel praktiline kogemus tervise näitajate modelleerimisel, mis toetab tervise dimensiooni lisamist mudelisse. Ülikool on vedanud Infotehnoloogilise Mobiilsusobservatooriumi (IMO) arendamist. Samuti toovad nad pardale kogemuse BIG-HEART terviseandmete modelleerimise vallas.

CentARi analüütikutel on samuti pikaajaline kogemus erinevate sotsiaalmajanduslike analüüside läbiviimisel ja mudelite ehitamisel. Lisaks täiendavad nad konsortsiumi tehnilise oskusteabega dünaamiliste mikrosimulatsioonimudelite arendamisel LIAM2 tarkvaras. CentAR on koostanud Rahandusministeeriumile LIAM2 põhise Pensionimudeli. Lisaks on CentARil kogemus ühiskonna protsessi modelleerimisel. Nende kogemuspagasisse kuulub näiteks projekti „Muutuvast tööturust ja rändest tulenevad võimalused ja väljakutsed haridus-, tööturu ja sotsiaalsüsteemis“ raames loodud rände tulu-kulumudel.

Konsortsiumil on olemas vajalik **tehniline taristu**. Projekti käigus kavandatakse tundlike registriandmete töötlemine korraldada läbi Statistikaameti turvalise keskkonna. See tagab andmete töötlemise tehnilise turvalisuse. Lisaks toetub konsortsium IMO andmestikele, mille kasutamises on Tartu Ülikoolil pikaajaline kogemus. BIG-HEART andmete töötlemiseks on konsortsiumil juurdepääs Tartu Ülikooli teadusarvutuskeskuse ressurssidele.

Tartu Ülikool ja CentAR ei tee **koostööd** esimest korda – näide varasemast koostöökogemusest suuremahulises teadusprojektis, mille raames arendati välja ka simulatsioonimudel, on juba eelpool viidatud projektist „Muutuvast tööturust ja rändest tulenevad võimalused ja väljakutsed haridus-, tööturu ja sotsiaalsüsteemis“. Käesolevas projektis toimib tööjaotus järgmiselt: Tartu Ülikool vastutab eeskätt rahvastikuprognoside ning tervise- ja regionaaldimensiooni sisu eest, CentAR keskendub eeskätt leibkonna mõõtmele ning mikrosimulatsiooni-spetsiifilise kompetentsi jagamisel ja rakendamisele. Samas töötatakse ühtse meeskonnana ning väiksemas mahus panustavad erinevad osapooled teineteise töösse.

Konsortsiumi liikmed on tugevad andmeanalüütikud, valdavad kaasaegseid statistikatarkvarasid ning oskavad integreerida erinevaid analüüsimeetodeid ja andmestikke.

Kokkuvõttes on konsortsiumil olemas nii teaduslikud teadmised kui ka tehnilised oskused, vajalik taristu ja varasem koostöökogemus, mis loob head eeldused projekti edukaks elluviimiseks.

4. Eelarve

Projekti eelarve on lisatud eraldi failina (Lisa 2. Eelarve ja tööde jaotus.xlsx). Seda koostades on silmas peetud, et mudel võib testimise käigus vajada täiendamist ning selleks tuleb ette näha nii aja, kui rahaline ressurss. Juhtimiskulude planeerimisel on lähtutud heast tavast, et need ei tohiks moodustada rohkem kui 10% eelarvest. Vt ka selgituse eelarve faili selgituste lehel.

5. Eetika

5.1. Teaduseetika teemade kontrollnimekiri

Teema	Kas on puutumus selle uuringuga?	Milles puutumus seisneb ja kuidas plaanitakse eetilisi riske maandada?
1. Inimesed: kas projekti raames läbiviidavasse uurimistöösse kaasatakse inimesi?	Ei	
2. ISIKUANDMED: kas uurimistöös töödeldakse isikuandmeid? (Märkige „jah“ ka juhul, kui isikuandmed uurimistöö käigus anonüümitakse.)	Jah	<p>Uurimistöö läbiviimiseks kaustatakse Infotehnoloogilise Mobiilsusobservatooriumi andmeid, millele lisatakse täiendavaid projekti jooksul täiendavaid andmeid teistest registritest. Lõplik andmevajadus selgub projekti käigus.</p> <p>Rahvaloenduste ja registrite andmebaaside kättesaadavaks tegemine ning kasutamine toimub Statistikaameti vahendusel, mis on IMO partneriks. Andmeid kasutatakse Statistikaameti turvalises andmetöötluskeskkonnas. Andmed on pseudonümiseeritud (otsest tuvastamist võimaldavaid tunnuseid nagu nimi ja isikukood andmestikus ei ole, on pseudokood, mille alusel erinevaid andmestikke ühendatakse).</p> <p>BIG-HEART andmeid analüüsib Taavi Tillmanni meeskond kel juba ligipääs andmetele läbi Tartu Ülikooli teadusarvutuse keskuse üliturvalise (SAPU) serveri.</p>
3. INIMESE EMBRÜONAALSED TÜVIRAKUD JA EMBRÜO: kas uurimistöös kasutatakse inimese embrüonaalseid tüvirakke või inimese embrüot?	Ei	
4. INIMESE KOED JA RAKUD: kas uurimistöös kasutatakse inimese rakke, kudesid või kehavedelikke?	Ei	
5. LOOMAD: kas uurimistöös kasutatakse loomi?	Ei	
6. GENEETRILISED RESSURSID ja nendega	Ei	

seotud teadmine: kas uurimistöös kasutatakse taimset, loomset (v.a. inimesed), mikroobset või muud päritolu geneetilisi ressursse või nendega seotud traditsioonilisi teadmisi?		
7. EUROOPA LIIDU VÄLISED RIIGID: kas uurimistöö toimub täielikult või osaliselt mõnes EL-i välises riigis või edastatakse isikuandmeid või muid uuringumaterjale sellisesse riiki?	Ei	
8. KESKKOND, TERVIS JA OHUTUS: kas uurimistöös plaanitavad tegevused või kasutatavad materjalid võivad kahjustada keskkonda või inimeste tervist?	Ei	
9. TEHISINTELLEKT: kas uurimistöö raames kasutatakse või arendatakse tehisintellekti süsteeme või meetodeid?	Ei	
10. TEADUSTULEMUSTE VÄÄRKASUTUS: kas uurimistöö raames luuakse materjale, meetodeid, tehnoloogiaid või teadmist, mida on võimalik kasutada ebaeetilistel või kahjulikel eesmärkidel?	Ei	
11. MUUD EETILISED TEEMAD: kas uurimistöö raames esineb muid eetilisi küsimusi, mida eelnevad küsimused ei käsitlenud? (Näiteks võib uurimisteema olla poliitiliselt tundlik või polariseeriv).	Ei	

5.2. Hea teadustava järgimine

5.2.1. Hea teadustava järgmise kinnitus

Uurimisrühm lähtub oma tegevuses Eesti teadlaste koodeksist ning rahvusvaheliselt tunnustatud hea teadustava põhimõtetest. Kõik uurimisrühma liikmed on kohustatud järgima akadeemilise aususe, sõltumatuse ja läbipaistvuse nõudeid nii teadustöö kavandamisel, läbiviimisel kui ka tulemuste avaldamisel. Projekti kõik etapid – alates andmete kogumisest ja analüüsist kuni tulemuste levitamiseni – dokumenteeritakse viisil, mis võimaldab töö reprodutseeritavust ja kontrollitavust.

Andmete kasutamisel järgitakse andmekaitse põhimõtteid. Tundlikke isikuandmeid töödeldakse pseudonüümistult või anonüümiseeritult, ligipääs andmetele on piiratud ning toimub kooskõlas seadusandluse ja andmekaitseasutuste juhistega.

Autorlus ja intellektuaalne omand jagunevad vastavalt panusele. Kõigi projekti väljundite puhul tuuakse selgelt välja kaasautorid, rahastaja ja institutsiooniline kuuluvus. Avalik suhtlus toimub vastutustundlikult: meedias ja sotsiaalmeedias jagatakse kontrollitud tulemusi, vältides eksitavaid üldistusi.

Uurimisrühm soodustab avatud teadust: võimalusel tehakse meetoodika, mudelid ja lähtekoodid avalikult kättesaadavaks, arvestades siiski ka andmekaitse nõuetega. Samuti edendatakse noorte teadlaste kaasamist ja juhendamist, pakkudes neile võimalust osaleda teadustöös.

Potentsiaalsetest huvide konfliktidest teavitatakse aegsasti ning otsuste tegemisel välditakse olukordi, mis võiksid seada kahtluse alla teadustöö sõltumatuse.

5.2.2. Kas projekti raames läbi viidav uuring või uuringud vajavad eetikakomitee kooskõlastust?

Jah

5.2.3. Eetikakomitee kooskõlastus(ed)

5.2.4. Kas geneetilisi ressursse käsitlevad teadusuuringud kuuluvad Nagoya protokolliga ja ELi ABS-määruse reguleerimisalasse?

Ei

5.3. Avatud teadus ja teadusandmete haldamine

5.3.1. Ülevaade teadusandmete haldamisest

Projektis kasutatavad andmed pärinevad peamiselt Eesti Statistikaameti hallatavatest ja andmekogudest. Andmete töötlemise ja analüüsi tegevused viiakse läbi Statistikaameti turvalises uurimiskeskonnas. Uurimisrühmal puudub otsene ligipääs isikuandmeid sisaldavatele andmefailidele väljaspool seda keskkonda.

Projekti käigus ei moodustu avalikult kättesaadavaid andmekogumeid. Kuna andmed sisaldavad isiku- ja majapidamistasandi tundlikku teavet, on juurdepääs neile piiratud ja seda saab taotleda ainult Statistikaameti vahendusel vastavalt kehtivatele reeglitele. Uurimistulemuste reprodutseeritavus tagatakse metoodika, koodide ja mudelite dokumenteerimise kaudu.

Projektis järgitakse isikuandme kaitse seadust ning Statistikaameti andmete kasutamise reegleid. Andmeanalüüsi koodid ja metoodilised kirjeldused on avalikud. Sel viisil tagatakse, et andmete haldus on kooskõlas hea teadustava ja isikuandmete kaitse nõuete.

5.3.2. Avatud teadusandmed: kas teadusandmed või osa neist tehakse kõigile tasuta kättesaadavaks ja vabalt kasutatavaks?

Ei. Meie projekti väljundid sisaldavad pseudonümiseeritud isikuandmeid, neid ei saa kõigile kättesaadavaks teha. Andmetele tekib juurdepääs, mis on sarnane juurdepääsuga IMO andmetele või juurdepääsuga pensionimudeli andmetele.

Ei. Kuna BIG-HEART andmebaas sisaldab suure hulga tundlike ja eriliiki isikuandmeid, siis ligipääs on piiratud vaid nimetatud põhjusel nimetatud teadlastele.

5.3.3. Piiratud juurdepääsuga andmed: kas projekti käigus luuakse teadusandmeid, mida ei saa avatult kättesaadavaks teha?

Jah. Kogu simulatsioonimudeli baasandmestik (lähtepunkt, kust simulatsioon käivitatakse) on piiratud juurdepääsuga, kuna see sisaldab pseudonümiseeritud isikuandmeid.

5.3.4. Jagatud intellektuaalomandi õigused: kas projekti raames loodud intellektuaalse omandi õigused jagunevad mitme teadusasutuse või juriidilise isiku vahel?

Projekti raames loodud intellektuaalne omand jaguneb kahe juriidilise isiku vahel – Tartu Ülikool, Eesti Rakendusuuringu Keskus CentAR – vastavalt panusele. Taotluse koostamisel oleme lähtunud sellest, et mõlemad organisatsioonid panustavad projekti võrdselt, seega jaguneb ka intellektuaalne omand vastavalt sellele.

Allikad

Andreassen, L., Fredriksen, D., Gjefsen, H.M., Halvorsen, E. and Stølen, N.M. (2020): [The dynamic cross-sectional microsimulation model MOSART](#). *International Journal of Microsimulation* 13 (1), 97-119.

Dekkers, Gijs. 2015. On the Modelling of Immigration and Emigration Using LIAM2. <https://doi.org/10.13140/RG.2.1.4373.8967>.

Jia, Zhiyang, Stefan Leknes, and Sturla A. Løkken. 2023. 'Moving beyond Expectations. From Cohort-Component to Microsimulation Projections'. Discussion Papers, Statistics Norway, Research Department, Oslo, No. 999.

- Järve, J., Anspal, S., Vörk, A., Leppik, L., Piirits, M. 2023. Lähtheraport. Rahandusministeeriumi riikliku pensionimudeli arendamine. CentAR.
- Keegan, M. 2007. Modelling the workers of tomorrow: the APPSIM dynamic microsimulation model. HILDA Survey Research Conference 2007.
- Li, Jinjing, and Cathal O'Donoghue. 2013. 'A Survey of Dynamic Microsimulation Models: Uses, Model Structure and Methodology'. *IJM* 6 (2): 3–55. <https://doi.org/10.34196/ijm.00082>.
- Lõo, L., Umova, N., Oja, M., Reisberg, S., Uusküla, A., Koldeb, R., Tillmann, T. (2025) Data Resource Profile: Linking electronic health and social records to study and lower health inequalities in cardiovascular diseases (BIG-HEART). doi: <https://doi.org/10.1101/2025.05.09.25327142>
- Morency, J-D., Vézina, S. and Dion, P. 2025. The Demosim microsimulation model at Statistics Canada: A tool for policy planning and evaluation. Statistics Canada. Catalogue no. 17-20-0001.
- Münnich, R., Schnell, R., Brenzel, H., Dieckmann, H., Dräger, S. Emmenegger, J., Höcker, P., Kopp, J., Merkle, H., Neufang, K., Obersneider, M., Reinhold, J., Schaller, J., Schmaus, S., Stein, P. 2021. A Population Based Regional Dynamic Microsimulation of Germany: The MikroSim Model. *methods, data, analyses* | 2021, Vol. 15(2), pp. 241-264
- Oja M, Tamm S, Mooses K, Pajusalu M, Talvik HA, Ott A, Laht M, Malk M, Lõo M, Holm J, Haug M. (2023) Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) common data model: lessons learned. *JAMIA open*. 2023 Dec 1;6(4):ooad100.
- Smith, S.K., Tayman, J. Swanson, D.A. A Practitioner's Guide to State and Local Population Projections. The Springer Series on Demographic Methods and Population Analysis 37. <https://doi.org/10.1007/978-94-007-7551-0>
- Spielauer, M. (2007). Dynamic microsimulation of health care demand, health care finance and the economic impact of health behaviours: Survey and review. *International Journal of Microsimulation*, 1(1), 35–53. <https://doi.org/10.34196/ijm.00005>
- Spielauer, Martin, Gerard Thomas Horvath, and Marian Fink. 2020. 'microWELT: A Dynamic Microsimulation Model for the Study of Welfare Transfer Flows in Ageing Societies from a Comparative Welfare State Perspective'. WIFO Working Papers.
- Vanella, Patrizio, Timon Hellwagner, and Philipp Deschermeier. 2023. 'Parsimonious Stochastic Forecasting of International and Internal Migration on the NUTS-3 Level – an Outlook of Regional Depopulation Trends in Germany'. *Vienna Yearbook of Population Research* 21 (April). <https://doi.org/10.1553/p-5pn2-fmn8>.
- Weymeirsch, Jan, Julian Ernst, and Ralf Münnich. 2024. 'Model Recalibration for Regional Bias Reduction in Dynamic Microsimulations'. *Mathematics* 12 (May):1550. <https://doi.org/10.3390/math12101550>.
- Zucchelli, E., Jones, A. M. & Rice, N. The evaluation of health policies through dynamic microsimulation methods. *International Journal of Microsimulation*. 5, 2–20 (2012). <https://microsimulation.pub/articles/00064>