NB! Print view does not include files that are attached to the form. Files should be printed out separately.

TEET431 application 2026

Research project application

Title in Estonian	EST-Health-30 - Eesti terviseandmete väärindamine
Title in English	EST-Health-30 – Valorisation of Estonian Health Data
Principal investigator	Jaak Vilo
Responsible institution	University of Tartu, Faculty of Science and Technology, Institute of Computer Science
Study start date	01.10.2024
Study end date	31.12.2027
Fields of research	1.2 Computer and information sciences 3.3 Health sciences

Main objective of the study (in Estonian)

Eesmärk on Eesti terviseandmete väärindamine läbi andmeteaduse meetodite, et pakkuda kvaliteetset tõendust tõhusate patsiendikesksete tervishoiu- ja ennetusteenuste osutamiseks.

Uuring viiakse läbi kahes etapis:
1. etapp on alusandmestiku loomine ja andmekvaliteedi tõstmise meetodite arendamine, mille käigus lingitakse tervise infosüsteemi, Tervisekassa andmekogu, retseptikeskuse, vähiregistri ja surma põhjuste registri andmed ja seeläbi koostatakse teadusuuringute läbiviimiseks alusandmestik . Alusandmestiku loomise oluliseks osaks on andmekvaliteeti ja andmehõivet parandavate meetodite sh tehisintellekti mudelite arendamine. Samuti on alusandmestiku loomine vajalik 2. etapi ellu viimiseks ja eesmärkide täitmiseks.

2. etapp on teadus- ja rakendusuuringute teostamine uurimismetoodikate arendamiseks Eesti terviseandmetel. Lähtume põhimõttest, et loodavad analüüsimeetoid on rakendatavad sõltumata uuringu all olevast haigusest või patsiendigrupist. Tulenevalt käimasolevatest teadus- ja rakendusuuringu projektidest keskendutakse taotletava teadusuuringu raames ravijärgimuse, ravitrajektooride ja -teekondade ning personaliseeritud ennetusmeetoditega seotud analüüsimeetodite arendamisele.

Main objective of the study (in English)

The aim is to valorise Estonian health data through data science methods in order to provide high-quality

evidence for the delivery of effective, patient-centred healthcare and preventive services.

The study will be conducted in two stages:

The first stage involves creating a foundational dataset and developing methods for improving data quality. Data from the Health Information System, the Estonian Health Insurance Fund database, the Prescription Centre, the Estonian Cancer Registry and the Causes of Death Registry will be linked to create a foundational dataset for conducting scientific research. An important part of creating the foundational dataset is the development of methods, including artificial intelligence models, to improve data quality and data capture. The creation of the foundational dataset is also necessary for implementing the second stage and achieving its objectives.

The second stage involves conducting scientific and applied research to develop research methodologies for Estonian health data. The study is based on the principle that the analysis methods developed will be applicable regardless of the disease or patient group under investigation. In view of the ongoing scientific and applied research projects, the proposed study will focus on developing analysis methods related to treatment adherence, treatment trajectories and pathways, and personalised prevention methods.

Other involved institutions

Additional information

Is this a student project?	No
Which subcommittee would you prefer to review your application?	3) Health Data and Data Science Subcommittee
Will part of the study be conducted outside Estonia?	No
Briefly explain the research activities taking place outside Estonia	
Has the study or any part of it been submitted for review to another ethics committee in Estonia or abroad?	Yes
Additional information about previous or concurrent reviews	EBIN: 11.11.2024 1.1-12/2793 23.03.2025 1.1-12/817

	20.12.2025 1.1-12/3216 17.03.2026 1.1-12/698
Has artificial intelligence been used in preparing the application or any of its parts?	No
Please explain for what purpose and to what extent AI was used	

Principal investigator

Person	Degree	Institution and occupation	Start date	End date	CV
Jaak Vilo	Doctor's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Professor of Bioinformatics (1,00), University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Head of Institute (0,10)	11.11.2024	31.12.2027	EST / ENG

Other researchers involved (investigator, assisting staff)

Role	Person	Degree	Institution and occupation	Start date	End date	CV
Investigator	Raivo Kolde	Doctor's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Associate Professor of Health Informatics (1,00)	11.11.2024	31.12.2027	EST / ENG
Investigator	Sven Laur	Doctor's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Associate Professor in Privacy-Preserving Data Mining (1,00)	11.11.2024	31.12.2027	EST / ENG
			University of Tartu,			

Role	Person	Degree	Institution and occupation	Start date	End date	CV
Investigator	Sulev Reisberg	Doctor's Degree	Faculty of Science and Technology, Institute of Computer Science, Associate Professor of Health Informatics (1,00)	11.11.2024	31.12.2027	EST / ENG
Investigator	Marek Oja	Doctor's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Research Fellow of Health Informatics (1,00)	11.11.2024	31.12.2027	EST / ENG
Investigator	Markus Haug	Master's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Specialist of Health Informatics (0,50)	11.11.2024	31.12.2027	EST / ENG
Investigator	Harry-Anton Talvik	Master's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Specialist of Health Informatics (0,50)	11.11.2024	31.12.2027	EST / ENG
Investigator	Hendrik Šuvalov	Master's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Specialist of Health Informatics (0,50)	11.11.2024	31.12.2027	EST / ENG
Investigator	Maria Malk		University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Junior Research Fellow in Health Informatics (0,75)	11.11.2024	21.12.2027	EST / ENG
Investigator	Ami Sild	Master's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Health Data Engineer-	11.11.2024	31.12.2027	EST / ENG

Role	Person	Degree	Institution and occupation	Start date	End date	CV
			Analyst (1,00)			
Investigator	Sirli Tamm	Master's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Senior Data Engineer (0,20)	11.11.2024	31.12.2027	EST / ENG
Investigator	Kerli Mooses	Doctor's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Research Fellow of Health Informatics (1,00)	11.11.2024	31.12.2027	EST / ENG
Investigator	Asad Munir	Doctor's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Lecturer in AI in Health Informatics (1,00)	23.03.2025	31.12.2027	EST / ENG
Investigator	Taavi Tillmann	Doctor's Degree	University of Tartu, Faculty of Medicine, Institute of Family Medicine and Public Health, Associate Professor of Public Health (1,00)	11.11.2024	31.12.2027	EST / ENG
Investigator	Õie Renata Siimon	Master's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Junior Research fellow in Health Informatics (0,50), University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Scientific Programmer (0,50)	11.11.2024	31.12.2027	EST / ENG
			University of Tartu,			

Role	Person	Degree	Institution and occupation	Start date	End date	CV
Investigator	Maarja Pajusalu	Master's Degree	Faculty of Science and Technology, Institute of Computer Science, Junior Research Fellow of Health Informatics (1,00)	11.11.2024	31.12.2027	EST / ENG
Investigator	Kermo Saarse	Master's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Junior Research Fellow in Health Informatics (1,00)	11.11.2024	31.12.2027	EST / ENG
Investigator	Helene Loorents		University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Specialist of Health Data (0,70)	11.11.2024	31.12.2027	EST / ENG
Investigator	Nikita Umov		University of Tartu, Faculty of Medicine, Institute of Family Medicine and Public Health, Junior Research Fellow in Public Health (1,00)	11.11.2024	31.12.2027	EST / ENG
Investigator	Laura Lõo	Master's Degree	University of Tartu, Faculty of Medicine, Institute of Family Medicine and Public Health, Junior Research Fellow of Public Health (1,00)	11.11.2024	31.12.2027	EST / ENG
Investigator	Neeme Ilves		University of Tartu, Faculty of Medicine, Institute of Clinical Medicine, Junior Research Fellow in Radiology (1,00)	11.11.2024	31.12.2027	EST / ENG
			University of Tartu,			

Role	Person	Degree	Institution and occupation	Start date	End date	CV
Investigator	Mirjam Jesmin	Master's Degree	Faculty of Science and Technology, Institute of Computer Science, Health Data Analyst (1,00)	20.12.2025	31.12.2027	EST / ENG
Investigator	Greete Kelli Aava	Master's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Data Engineer (1,00)	17.03.2026	31.12.2027	EST
Investigator	Belinda Lepmets	Master's Degree		01.09.2026	31.12.2027	EST / ENG
Investigator	Mart-Mihkel Aun	Master's Degree		01.09.2026	31.12.2027	EST / ENG
Investigator	Anna Zadorožnaja	Master's Degree		01.09.2026	31.12.2027	EST / ENG
Investigator	Fareeda Madeen Gouher Mohammad	magistrikraad	nooremteadur	01.09.2026	31.12.2027	
Investigator	Hendrik Aruoja			01.09.2026	31.12.2027	EST / ENG
Investigator	Serle Kirsipuu			20.12.2025	31.12.2027	EST / ENG
Investigator	Christine Kaldoja			20.12.2025	31.12.2027	EST / ENG
Assisting staff	Kaia Kask	Doctor's Degree	University of Tartu, Faculty of Science and Technology, Institute of Computer Science, Assistant (1,00)	01.09.2026	31.12.2027	EST / ENG

Role	Person	Degree	Institution and occupation	Start date	End date	CV
Explanation of the researchers' roles and responsibilities	Vastutav(ad) uurija(ad) juhib(vad) uuringut, koordineerib kaasuurijate tööd ning vastutab nõuete järgimise eest. Teised uurijad osalevad vastavalt oma pädevusele andmete ettevalmistamises ja analüüsimises, tulemuste tõlgendamises ning dokumenteerimises. Andmetele pääsevad ligi ainult tööülesannete täitmiseks vajalikud uurijad. Kaastöötajal andmetele ligipääsu ei ole, tema tegeleb eetika taotlustega, otsustega ja nende koordineerimisega.					

3.1. Funding

Funding sources [311]

● SA Eesti Teadusagentuuri rahastatud Personaalse uurimistoetuse rühmagrant "Kliiniliselt oluliste radade tuvastamine terviseandmetest" (PRG1844, jaan 2023 – dets 2027). Vastutav täitja: Jaak Vilo. Eelarve:
1 080 000 EUR.
● SA Eesti Teadusagentuuri rahastatud grant SLTAT24173 (TK213U6) "Eesti Tehisintellekti Tippkeskus" (EXAI) (terviseinformaatika alamprojekt) (01.01.2024−31.12.2030); Vastutav täitja: Jaak Vilo; Eelarve: 525 222 EUR.
● SA Eesti Teadusagentuuri rahastatud temaatilise teadus- ja arendusprogrammi projekt "Terviseandmete teisese kasutamise võimekuse kasvatamine" (TEM-TA72, jaan 2024 - dets 2028). Vastutav täitja: Raivo Kolde. Eelarve:
1 146 150 EUR.
● Teadusarvutuste ja andmete säilitamiseks kasutatakse taristut Eesti Teadusarvutuste Infrastruktuur (ETAIS), mida toetab Haridus- ja Teadusministeerium otse Eesti Teadusarvutuste Infrastruktuuri ja Eesti Hariduse ja Teaduse Andmesidevõrgu kaudu1. ETAIS tegevustoetuse maht on 305 000 EUR aastas.

Estimated total cost of the study [312]

1 800 000

Related projects [313]

PRG1844 "Discovery and Analysis of Clinical Pathways in Health Data" (01.01.2023−31.12.2027); Principal Investigator: Jaak Vilo; University of Tartu, Faculty of Science and Technology, Institute of Computer Science; Financier: Estonian Research Council; Financing: 1 080 000 EUR.

SLTAT24173 (TK213U6) " Estonian Centre of Excellence in Artificial Intelligence" (01.01.2024−31.12.2030); Principal Investigator: Jaak Vilo; University of Tartu, Faculty of Science and Technology, Institute of Computer Science (partner); Financier: Ministry of Education and Research; Financing: 525 222 EUR.

TEM-TA72 "Enhancing the Capability of Secondary Usage of Health Data (TAKS)" (01.01.2024−31.12.2028); Principal

Investigator: Raivo Kolde; University of Tartu, Faculty of Science and Technology, Institute of Computer Science; Financier: Estonian Research Council; Financing: 1 146 150 EUR.

3.2. Compensation and insurance for participants

Will participants receive compensation (monetary and/or non-monetary) for taking part in the study? [321]	No
Amount of compensation and/or description and justification of non-monetary compensation [321a]	
Is insurance coverage for participants required for conducting the study? [322]	No
Explanation of insurance coverage [322a]	

4.1. Background, purpose, and justification of the study

Brief overview of previous studies on the same topic [411]	Avatud teaduskoostöö organisatsiooni OHDSI loodud ja edasiarendatavate OMOP (Observational Medical Outcomes Partnership) andmemudelil põhinevate rahvusvaheliste terviseuuringute arv on viimastel aastatel hüppeliselt kasvanud. Neid on avaldatud väga erinevatel teemadel - näiteks on uuritud nii rasvumist, põlveliigese vahetust, reumatoidartriiti kui ka koroonapatsiente[1,2,3,4]. Ka käesoleva uuringu meeskonnal on pikaaegne kogemus sarnaste OMOP andmemudelil põhinevate terviseuuringute läbiviimisel. Näiteks EHDEN projektis osaleb meeskond aktiivse partnerina OMOP mudeli ja analüüsimeetodite arendamisel ning nende populariseerijana Euroopas[5,6]. Projektis “Masinõppe ja AI toega teenused” lõime käesoleva taotluse raames loodava andmestikuga sarnase, kuid väiksemas mahus (valimi suurus 10%, ei sisaldanud andmeid

vähiregistrist, surma põhjuste registrist) kvaliteetse andmestiku 2012.-2019. aasta terviseandmetel[7,8,9,10]. Lisaks varasemalt loodud andmestikku ei uuendatud. Sellest alates oleme taaskasutanud ja edasi arendanud Eesti terviseandmete jaoks spetsiifilisi automaatseid andmepuhastuse tehnikaid, mis on võimaldanud meil endal läbi viia uuringuid ja ka osaleda mitmesugustes rahvusvahelistes projektides. Heaks näiteks on uuring, kus hindasime inimese papilloomiviiruse (HPV) tüüpide levimust Eestis tuginedes tekstikaevega tervisedokumentidest eraldatud HPV testi tulemustele[8]. Meie senise töö ja Eesti terviseandmete kvaliteeti näitab ka see, et DARWIN projektis valis Euroopa Ravimiamet meid üheks kvaliteetseks andmepartneriks, mille tulemusel kasutatakse Eesti OMOP kujule viidud Eesti Geenivaramu terviseandmestikku ühena teiste seas ravimite reguleerimisega seotud küsimuste uurimiseks[11]. OPTIMA projektis analüüsime Eesti terviseandmeid organkasvajatega patsientidele optimaalse ravi leidmiseks. Seniste uuringute tulemusi on avaldatud nii Eesti kui rahvusvahelistes teadusajakirjades, postritel kui ettekannetel, samuti uurimisgrupi töös osalenud üliõpilaste lõputöödes.

Study objective, summary, and justification [412]

Meditsiinis ja rahvatervises on otsuste tegemiseks vaja laiapõhjalistel teadusuuringutel põhinevat tõendust. Uuringute tarvis on reeglina vaja esinduslikke andmeid, mis on võimalikult värsked ja samal ajal piisavalt pika aegreaga. Näiteks on värskeimad andmed (sh imikute ja vastsündinute kohta) vajalikud vastamaks küsimustele, mis puudutavad terviseteenuse osutamise hetkeseisu, uute ravimeetodite kasutust ja efektiivsust ning tervisesüsteemi vastust erinevatele hiljutistele sündmustele. Usaldusväärsemate ennustusmudelite loomiseks, haiguse arengu, ravi efektiivsuse, erinevate ravitrajektooride jms hindamiseks on vajalik vaadelda pikemaid ajaperioode.

Ajakohastel andmetel tugineva tõenduse leidmiseks kasutatakse järjest enam päriselu terviseandmete teisesel kasutusel põhinevaid uuringuid. Tänu Eesti unikaalsele tervishoiuandmete korraldusele on võimalik ühendada patsiendi tervist puudutavad olulised komponendid haiguste (Tervise infosüsteem, vähiregister), ravimite (retseptikeskus), tarbitud teenuste ja kindlustuskaitse (Tervisekassa andmekogu) ning surma põhjuste kohta. Nende andmetel OMOP andmemudelil põhinevate terviseuuringute läbiviimisel on Eesti tervishoiuvaldkonnale ja ühiskonnale laiemalt mitmekordne kasu. Nii aitab igapäevase tervishoiuvaldkonna toimimise raames salvestatud andmete taaskasutamine ühelt poolt hoida

kokku andmete kogumisele kuluvat aega ja raha ning teisalt annab kõige täpsema ülevaate tegelikest protsessidest ja trajektooridest erineva terviseseisundi, soo ja vanusega isikutel. Oluline on siinjuures, et terviseandmete teisese kasutamise korral ei suurene patsientide uuringu koormus kuna kasutatakse varasemalt kogutud retrospektiivseid andmeid.

Terviseandmete teisene kasutamine võimaldab ilma lisakoormust põhjustamata analüüsida ja osutada efektiivsemalt tervishoiuteenuseid patsiendi gruppidel, kellel traditsiooniliste teadusuuringute tegemine ei ole eetilistel kaalutlustel võimalik (näiteks ravimite mõju lastele või rasedatele[1]). Kaasates analüüsi infot patsiendi kogu raviajaloo kohta pikema aja vältel, mitte ainult uuritava haigusega seotud tegevuste kohta, võib-olla võtmetähtsusega oluliste seoste avastamisel. Lisaks on taotletava teadusuuringu käigus 30% valimi peal välja töötatud meetodid andmekvaliteedi parandamiseks rakendatavad ka teistes teadusuuringutes ning riiklikes andmebaasides kogu rahvastiku andmetel. Taotletav teadusuuring loob eeldused, et päriselu terviseandmeid saaks Eestis laiemalt kasutada ning aitab parandada tervishoiuteenuste kvaliteeti, tervishoiusüsteemi efektiivsust ning toetada tervishoiupoliitiliste tõenduspõhiste otsuste tegemist, toetades seeläbi rahvastiku tervise arengukava 2020-2030 elluviimist. Kvaliteetsem tervishoiusüsteem omakorda toetab tervelt elatud eluaastate suurenemist.

Taotletav teadusuuring koosneb kahest etapist:
1. etapp: alusandmestiku loomine ja andmekvaliteedi tõstmise meetodite arendamine

Alusandmestiku loomisel pannakse kokku ja ühtlustatakse tervise infosüsteemist, Tervisekassa andmekogust, retseptikeskusest, vähiregistrist ja surma põhjuste registrist pärit andmed. Erinevad andmebaasid kajastavad erinevaid aspekte sama patsiendi raviteekonnast. Seetõttu on võimalik saada tervikpilt raviprotsessist vaid kõikide vajalike andmete ühendamise teel. Ülekatted andmeallikate vahel võimaldavad ka süstemaatiliselt hinnata algandmete kvaliteeti ja -hõivet, mis paljuski sõltuvad nende kogumise eesmärgist ning tehnilisest võimekusest ja digitaliseerituse astmest andmete kogumise hetkel. Näiteks saame alusandmestiku loomisel hinnata kui paljude patsientide kohta on olemas vajalik informatsioon, kui suur osa tehtud mõõtmistest on andmetes kajastatud, kas ja kuidas on puuduvad mõõtmised kallutatud või kui usaldusväärsed on diagnoosid. Kuna valim on piisavalt suur ja

esinduslik, siis saadud tulemused üldistuvad kogu Eesti terviseandmetele.

Tuginedes oma varasemale teadustööle analoogsete OMOP andmebaaside loomisel teame, et on mitmeid viise kuidas andmehõivet ja -kvaliteeti läbi erinevate andmepuhastamise protseduuride tõsta. Näiteks enne andmete analüüsimist teostame täiendava anonümiseerimise vabatekstilistes andmeväljades, et eemaldada võimalikud nimed, telefoninumbrid, isikukoodid, aadressid jms. Sellega vähendame isikute tuvastamise riski.

Teadusuuringu käigus on plaanis juba olemasolevat anünomiseerjat veelgi tõhustada. Lisaks on üheks väljakutseks tervise infosüsteemi andmete kasutamisel vabatekstilistes andmeväljades oleva olulise terviseinfo eraldamine, struktureerimine ja standardimine. Vabatekstilistes väljades andmeanalüüsiks oluline informatsioon võib esineda lihtsamini eraldataval kujul, nagu näiteks vererõhu, kehakaalu ja pikkuse mõõtmised, kui ka keerukamal kujul, nagu patsiendi kaebused või harjumuste kirjeldused, kus on oluline mõista ka teksti sisu. Selleks, et veelgi rohkem ja efektiivsemalt leida üles erinevate uurimisküsimustega seotud infot vaba tekstist on oluline arendada vastavaid meetodeid, sh tehisintellekti ja tekstikaeve meetodeid, keerukamate olukordade jaoks, mis vajaliku informatsiooni eraldaks ning viiks selle struktureeritud ja standardiseeritud kujule.

Töötades järjepidevalt samal andmestikul, on võimalik erinevate uurimisküsimuste lahendamisega andmekvaliteeti märgatavalt parandada. Töö käigus välja töötatud metoodikad loovad eeldused järgnevate uuringute paremaks läbiviimiseks ning on hiljem rakendatavad ka tervise infosüsteemis olevate andmete parandamiseks. Loodud metoodikate rakendamine toetab ka Eesti teadus- ja arendustegevuse, innovatsiooni ning ettevõtluse (TAIE) arengukava 2021-2035, kus üheks juhtmõtteks on liikumine teksti- ja dokumendipõhiselt terviseandme mudelilt üle andmepõhisele mudelile. Taotletavas teadusuuringus loodavaid andmepuhastustehnikaid on võimalik kasutada ka uue põlvkonna tervise infosüsteemi loomisel ja arendamisel.

Andmete esitus algandmeallikates on läbi aja muutunud ja arenenud. Näiteks Tervisekassa tervishoiuteenuste loetelu uuendatakse iga kolme kuu järel ning pikema aegrea puhul ei pruugi erinevatest laboritest pärinevad mõõtmised olla

samamoodi kodeeritud. Et uuringud saaks läbi viia võimalikult pikas ajaaknas, tuleb andmete kodeering viia ühtsele standardile. Selle protsessi käigus tuleb luua üleminekutabelid, kus algallikates esinevad koodid on esitatud standardselt. Nende loomisel kombineeritakse tekstikaeve, klasterdamise ja automaattõlke vahendeid manuaalse inspektsiooni ning kliinilise ekspertiisiga. Tulemuseks saadud üleminekutabelid võimaldavad automatiseerida andmete standardkujule viimist nii antud andmestiku uuendamisel kui ka täiesti uutes uuringutes. Samas on andmehõive ja -kvaliteedi parandamine iteratiivne protsess, kuna erinevate haigustega seotud uuringutes osalemisel tulevad ilmsiks uued kitsaskohad. Seetõttu on andmekvaliteedi tõstmise juures oluline andmeid kasutada võimalikult paljudes ja erinevates uuringutes.

Linkimise, puhastamise ning eelnevalt toodud lähenemiste rakendamisel viime erinevatest allikatest pärit andmed ühtsele OMOP andmemudelile, kus iga kirje on rahvusvaheliselt defineeritud standardkujul. Andmete OMOP kujule viimine teeb võimalikuks nende analüüsi juba suure hulga olemasolevate analüüsitööriistadega, mis muu hulgas võimaldavad keerukate kohortide loomist ning jagamist, patsiendigruppide kirjeldamist, ennustusmudelite treenimist ja palju muud. Nii saab uuringuid läbi viia ka hajutatult ehk täiendavate uurimisküsimuste lahendamisel (millele küsitakse eraldi eetikakomitee ja andmeandjate nõusolek) ei anta uutele isikutele ligipääsu andmetele vaid algne uuringumeeskond jooksutab läbi uuringu koodi ning edastab vaid agregeeritud tulemused. 2. etapi uuringu läbiviijad saavad andmetele ligipääsu vajaduspõhiselt jälgides minimaalsusprintsiipi läbi Eesti Teadusarvutuste infrastruktuuri hallatava turvalise töötlemiskeskkonna, kus kõikide kasutajate kõik tegevused logitakse.

Eesti riiklike terviseandmebaaside omanikud on teinud avalikult kättesaadavaks erinevat riikliku tervisestatistikat, mis annab mõningase ülevaate valitud tervisesündmuste toimumistest ja tervishoiuteenuste tarbimisest. Näitek TAI tervisestatistika ja terviseuuringute andmebaasis (statistika.tai.ee) on esmashaigestumuse info valitud RHK-10 peatükkide ja alampeatükkide kohta avaldatud kuni 2016 aastani ning hilisema perioodi avalik statistika puudub.

Uuringute paremaks planeerimiseks ja teostatavuse hindamiseks on vaja detailsemat infot, kus on kombineeritud mitme andmepunkti ja andmeallikat ning võetud arvesse sündmuste toimumise omavahelisi ajalisi seoseid. Näiteks

südameveresoonkonna haiguste ennetamiseks on oluline teada üle 40 aasta vanuste patsientide hulka, kellel on kolesterooli tase üle normi, kuid statiinipõhiseid ravimeid pole välja kirjutatud. Selleks on vaja andmeid labori analüüsidest (Tervise Infosüsteem) ja retsepikeskusest (Tervisekassa). Adresseerimaks seda puudujääki, plaanime põhjalikult kirjeldada loodud andmestikku läbi erinevate numbriliste näitajate ja teostatavusanalüüsi detailsusastmel säilitades k ≥ 5 anonüümsus. Selline ülevaade esinduslikust populatsioonipõhisest andmekogumist on väärtuslikuks vahendiks tervishoiupoliitika kujundajatele ja -teenuse osutajatele ning uute uuringute planeerimisel ja teostatavuse hindamisel.

1. etapi eesmärgid kokkuvõtvalt on:
Hinnata valimi põhjal Eesti terviseandmete kvaliteeti ja hõivet võrreldes erinevatest andmeallikatest pärinevat informatsiooni.
Andmekvaliteedi ja andmehõive tõstmise meetodite sh tehisintellekti-, tekstikaeve- ja imputatsioonimeetodite arendamine.
Tervisesündmuste esituse ühtlustamine rahvusvahelistele standarditele üle aja ja andmeallikate, sh kasutades tekstikaeve, klasterdamise ja automaattõlke meetodeid
Arendatavad meetodid on hiljem rakendatavad ka uutes uuringutes, suurematel andmehulkadel ning riiklikes infosüsteemides andmekvaliteedi ja -hõive parandamiseks.
Standardiseeritud ja ühtlustatud andmestiku loomine ja kirjeldamine teadusuuringute läbiviimiseks. Andmestikku kirjeldatakse läbi erinevate numbriliste näitajate ja teostatavusanalüüside senisest suuremal detailsusastmel.

2. etapi eesmärgid kokkuvõtvalt on:

Arendada haigustrajektooride ja raviteekondade analüüsi metoodikaid kasutades sh tehisintellekti meetodeid, et kirjeldada praeguseid ravipraktikaid, võrrelda neid raviteekondade ja -juhenditega ning modelleerida tervisetulemeid ja ravi majanduslikku mõju.

Analüüsida erinevate faktorite (nt sugu, vanus, erinevad haigused) mõju ravijärgimusele, hinnata ravijärgimuse mõju erinevatele tervisetulemitele ja luua personaalseid ennustusmudeleid, mis võimaldavad maandada ravi mittejärgimise riske.

Arendada personaliseeritud ennetusmudeleid, mis võimaldavad

vaatlusandmete põhjal tuvastada praeguste haiguse ennetuste ja ravipraktikate kitsaskohti, planeerida senisest paremini haiguste ennetustegevusi ja hinnata pakutud ennetusteenuste majanduslikku mõju

4.2. Sample and research methodology

Description of the sample [421]

Andmetiku aluseks on 30% juhuvalim Eestis alates 2012. aastast tervishoiuteenuseid tarbinud Eesti elanikest. Valim sisaldab kontrollgruppi.

Andmete vaatlemine üle pika ajaperioodi võimaldab analüüsida tervisesündmuste esituse ja kvaliteedi arengut läbi aja ning uurida terviklikke haigustrajektoore alates ennetustegevustest, esmasdiagnoosist, raviteenuste osutamisest lõpptulemini.

Data collection [422]

Andmetiku aluseks on 30% juhuvalim Eestis alates 2012. aastast tervishoiuteenuseid tarbinud Eesti elanikest (andmesubjektid). Nende andmesubjektide kohta saadakse pseudonüümitud andmed alates 2012 kuni 2026 aastani Tervisekassa andmekogust, retseptikeskuse andmekogust, Tervise Infosüsteemist, surma põhjuste registrist ja vähiregistrist.

Esimene väljastusel toimub 2024. aasta oktoobris, kui uurimisgrupile väljastatakse räsi abil pseudonüümitud andmed 2012-2023 aastate kohta. Depseudonüümimise võtit ehk räsifunktsioonis kasutatavat räsiparooli uurimisgrupile ei väljastata. Andmeid uuendatakse regulaarselt, kuid mitte tihedamini kui kord aastas, et mitte üle koormata Eesti tervishoiuandmete eest vastutavate asutuste nagu TEHIK ja Tervisekassa andmeväljastuse võimekust. Andmete uuendamine toimub järgmiselt:

2024 aasta andmed väljastatakse andmed 2025. aasta I kvartalis
2025 aasta andmed väljastatakse andmed 2026. aasta I kvartalis
2026 aasta andmed väljastatakse andmed 2027. aasta I kvartalis

Täpne väljastatav andmekoosseis on toodud taotluse lisas 1. Andmekoosseisu defineerimisel lähtusime eelnevate uuringute kogemusest ning kaasasime tunnused, mis on tüüpiliselt vajalikud andmete OMOP kujule viimiseks ja erinevate

terviseuuringute läbiviimiseks. Andmete OMOP kujule viimiseks ja puhastamiseks kasutatakse varasemates projektides loodud tarkvara ja töövoogusid, mida projekti raames täiendatakse vastavalt vajadusele. See tagab projekti kiire käivitamise ning iga-aastase ladusa andmete uuendamise.

Andmeid uuendatakse samamoodi nagu toimus algne andmete edastamine ehk kasutatakse sama räsifunktsiooni ja räsiparooli. Andmete uuendamise käigus isikute arv valimis suureneb, kuna andmestikku lisanduvad sünnid ja isikud, kellel varasemalt isikukood puudus. Nende andmete lisamine on oluline, sest see aitab tagada, et igal uuringu aastal on valimis 30% rahvastikust. Uute sündide lisamine võimaldab hinnata imikute ja väikelastega teostatud ravipraktikate muutust ajas ning seeläbi panustada haavatavate rühmade raviteenuste kvaliteedi parandamisse kasutades neid vähekoormavaid uurimismeetodeid.

Käesolevas uuringus uusi isikuandmeid ei koguta ja andmesubjektidega ühendust ei võeta. Kasutatakse juba olemasolevaid terviseandmeid teadusuuringute läbiviimise eesmärgil.

Data analysis [423]

Andmeid analüüsitakse retrospektiivselt. Andmekvaliteedi ja andmehõive hindamiseks võrreldakse eri andmeallikatest pärinevat teavet ning kasutatakse numbrilisi näitajaid ja teostatavusanalüüse. Andmekvaliteedi parandamiseks ja tervisesündmuste ühtlustamiseks kasutatakse tehisintellekti-, tekstikaeve-, imputatsiooni-, klasterdamise ja automaattõlke meetodeid. Ravitrajektooride, raviteekondade ja ravijärgimuse analüüsimisel kirjeldatakse ravipraktikaid, võrreldakse neid raviteekondade ja -juhenditega, hinnatakse erinevate tegurite mõju ravijärgimusele ja ravijärgimuse mõju tervisetulemitele ning koostatakse ennustusmudeleid. Samuti modelleeritakse tervisetulemeid ning ravi ja ennetusteenuste majanduslikku mõju. Analüüsiks kasutatakse Atlas serverit, Usagit, Data Characterisationi ning PopulationLevelEstimationi ja PatientLevelEstimationi R-pakette ning uurimisrühma arendatud tööriistu. Andmeanalüüs toimub turvalises SAPU serveris.

5.1. Recruitment and informing of participants

Are people directly involved as participants in the study? [511]	No
Recruitment of participants [511a]	
Justification for including vulnerable participants [512a]	

5.2. Impact of the study on participants

Procedures and activities with participants [521]	
Description of physical intervention [522a]	
Description of mental strain or risk of mental health impairment [523a]	

5.3. Use of human biological material

Is human biological material collected, used, or stored during the study? [531]	No
Description of collection and use of biological material [531a]	

5.4. Use of artificial intelligence

Is artificial intelligence used in the study? [541]	Yes
Description of artificial intelligence use [541a]	Alusandmestiku loomise oluliseks osaks on andmekvaliteeti ja andmehõivet parandavate meetodite sh tehisintellekti mudelite arendamine. Keerukamate kontseptsioonide eraldamiseks tuleb

arendada tehisintellekti meetodeid nagu BERT ja keelemudelid, mis suudavad eraldada fakte, mille puhul on oluline teksti kontekst ja mõttest aru saamine. Haigustrajektooride ja raviteekondade modelleerimisel on mitmeid väljakutseid, mis tulenevad peamiselt kliiniliste juhtumite keerukusest ning patsientide multimorbiidsusest.

Andmetiku aluseks on 30% juhuvalim Eestis alates 2012. aastast tervishoiuteenuseid tarbinud Eesti elanikest. Nende andmesubjektide kohta saadakse pseudonüümitud andmed alates 2012 kuni 2026 aastani Tervisekassa andmekogust, retseptikeskuse andmekogust, Tervise Infosüsteemist, surma põhjuste registrist ja vähiregistrist. Juhuvalimi tõttu on kõikidel inimestel võrdne tõenäosus valimisse sattuda.

Käesolevas uuringus uusi isikuandmeid ei koguta ja andmesubjektidega ühendust ei võeta. Kõikide uuringute puhul kaasame ka kliinilised eksperdid, kes aitavad tulemusi konteksti panna ning tagavad uuringutulemuste rakendatavuse.

Andmete edasisel töötlemisel SAPUs ja uuringu eesmärkide täitmiseks kasutatakse OMOP mudeli jaoks loodud ja publitseeritud tööriistu (Atlas server, Usagi, Data Characterisation, PopulationLevelEstimation ja PatientLevelEstimation R-paketid, uurimisrühma arendatud tööriistad jt). Uurimisgrupp tagab, et andmetele ligipääs on vastavuses kehtiva seadusandlusega. Vastutavad uurijad kontrollivad enne tulemuste tagastamist, et saadud tulemused on kooskõlas eetikakomitee antud loaga ja järgivad k ≥ 5 anonüümsuse printsiipi.

5.5. Ethical risks and their mitigation

Respect for participants [551]

Uuritavate isikuandmeid töödeldakse isikuandmete kaitse üldmääruse (EL) 2016/679 artikkel 9(2)(j) ja isikuandmete kaitse seaduse (IKS) § 6 lõigete 1, 3 ja 4 alusel teadusuuringu läbiviimise eesmärgil. Isikuandmete töötlemise õigusliku aluse põhjalikum hinnang on toodud käesoleva taotluse lisas 2 “EST-Health-30 andmekaitsealane mõjuhinnang”, peatükis “töötlemise õiguslikud alused”.

Võrdse kohtlemise printsiip – käesolev projekt ei diskrimineeri kedagi rahvuse, soo, vanuse ega muu kuuluvuse alusel. Kõikide ühiskonna- ja vanusegruppide puhul on eesmärk toetada elukvaliteedi parandamist ja tervena elatud aastate arvu suurendamist. Juhuvalimi tõttu on kõikidel inimestel võrdne tõenäosus valimisse sattuda.

Benefits and harms associated with the study [552]

Heategemise printsiip – antud projektist saadav kasu on uuritavatele pigem kaudne, kuna alustatakse baasteadmiste loomisega ja otsene kliinilises praktikas kasutatav kasu haigusega tegelemiseks võib tekkida alles mitme aasta pärast. Seega saavad antud tööst kasu pigem tulevased patsiendid.

Kahju vältimise printsiip - Planeeritav projekt on kooskõlas mittekahjustamise printsiibiga, kuna uuring ei koorma patsiente ega põhjusta riske nende tervisele.

Other ethical risks [553]

Ei.

6.1. Personal data processing control questions

Are personal data processed in the study? [611]

Yes

Explanation for not processing personal data [611a]

6.2. Origin of personal data

Are personal data processed that are not collected directly from the individual? [621]

Yes

Are personal data processed from national databases or registries? [622]

Yes

Which national databases are used in the study? [623]

- [ ] 1) gene bank
- [x] 2) health information system
- [x] 3) cancer registry
- [ ] 4) cancer screening registry
- [ ] 5) pregnancy information system
- [ ] 6) myocardial infarction registry
- [ ] 7) tuberculosis registry
- [ ] 8) drug treatment database

☑ 9) causes of death registry
☐ 10) Infectious diseases registry
☑ 11) prescription center
☑ 12) health insurance fund database
☐ 13) other

Explanation regarding the use of data from databases [623]

Eesti eri terviseandmekogudes sisaldub patsientide tervise kohta erinevat liiki ja eri kvaliteediga infot. Selleks, et saada tervisevaldkonna küsimuste uurimiseks patsiendi terviseseisundist kvaliteetne pilt, on vajalik vaadelda korraga andmeid mitmest andmekogust.

Tervise infosüsteemi epikriisides on kirjas diagnoosid, raviskeemid, saatekirjades ja saatekirja vastustes ka laborianalüüside tulemused, kuid info võib olla puudulik. Epikriisides sisaldub vaid haigusloo kokkuvõte, mitte aga kõik teostatud uuringud või väljakirjutatud ravimid, mis mitmetes uuringutes on oluline info. Teiselt poolt on epikriisid ainsad tervisedokumendid, kus sisaldub vabas vormis kirja pandud info patsiendi kaebuste, üldseisundi, allergiate, ravi kõrvalmõjude jms kohta. Tekstilistest osadest suudame eraldada tehisintellekti meetodite abil ka muud olulist infot, mis kodeeritud väljades puudub, näiteks patsiendi kaebusi, ravimite nõrgemaid kõrvalmõjusid. Tervisekassa andmekogu info tervishoiuteenuste kohta on täielikum ja sisaldab ka teenuste hindu, kuid selle detailsus on samas madalam (mitmed tervishoiuteenused märgitud sama koodiga) ja laborianalüüside kohta puuduvad analüüsitulemused. Kõige parema pildi patsiendile välja kirjutatud ravimite osas annab retseptikeskus, lisaks on seal ka info ravimi väljaostmise kohta, mis on väga oluline indikaator ravijärgimuse hindamiseks. Samas puudub retseptikeskuses info käsimüügiravimite kohta (seda infot võib potentsiaalselt leida epikriisidest). Paljude tervisevaldkonna uuringute puhul (nt pahaloomulised kasvajad) on oluline uurida suremust, selleks on kõige kvaliteetsem info kirjas surma põhjuste registris, mis sisaldab nii surma kuupäeva kui ka spetsialisti poolt kinnitatud surma põhjuseid.

Andmeväljade valikul oleme lähtunud uurimistöö eesmärkidest tuginedes uurimismeeskonna senistele kogemustele ja minimaalsuse printsiibist. Laiapõhjalisus võimaldab hinnata andmekvaliteeti üle paljude haiguste, sh arvestades kaasuvaid haigusi, ja tagab, et loodud andmetöötlusmeetodid üldistuvad uutele uuringutele sõltumata vaatluse all olevast haigusest ja patsiendi gruppidest. Lähtuvalt sellest, et üheks taotletava

teadusuuringu eesmärgiks on, et arendatavad analüüsimeetodid ja ennetusmeetodid üldistuksid ka uutele andmestikele ja haigustele, ei saa me piirata andmestikku ei diagnooside, vanuse ega muude parameetrite järgi. Iga piirang vähendaks andmestiku esinduslikkust üldpopulatsiooni suhtes ja muudaks arendatavad meetodid vähem üldistuvaks ning piiraks loodud meetodite kasutamist teistes uuringutes.

Ajaperioodi määratlemisel oleme lähtunud varasemast kogemusest 2012.-2019. aasta andmetega ning teadustöö eesmärkidest. Pikk ajaperiood tagab, et andmetes on näha tervisesündmuste esituse ja kvaliteedi areng läbi aja ning me saame uurida terviklikke haigustrajektoore, alates ennetustegevustest, esmasdiagnoosi, raviteenuste osutamise ja lõpptulemini. Samuti võimaldab loodava andmestiku ajaraam hinnata riskimudelite 10-aastast ennustusvõimet vastavalt rahvusvahelistelt levinud praktikatele.

Have you previously consulted the data controller on the availability of the data and the feasibility of conducting the study? [623b]

Yes

Please explain why you did not consult the data controller regarding the study. [623c]

Please provide a document or a copy of correspondence confirming that the data controller is aware of the planned study and that the requested data can be provided for the purposes of the study

How is the secure processing of personal data from databases ensured? [624]

Andmeandjatelt saadud algandmed paigaldatakse SAPUs spetsiaalselt kirjutuskaitstud kausta, et vältida andmete juhuslikku hävimist või kahjustumist. SAPU-st tehakse regulaarselt krüpteeritud varukoopiaid Eesti Teadusarvutuste infrastruktuuri lindirobotile, mis asub füüsiliselt teises asukohas (andmekeskuses).

Andmed arhiveeritakse üks aasta peale projekti lõppu. Andmete arhiveerimine üheks aastaks peale projekti lõppu on vajalik

retsenseerimisel olevate artiklite publitseerimisprotsessi lõpetamiseks. Arhiveerimisperioodi lõpus, hiljemalt 31.12.2028 andmed kustutatakse.

Projekti käigus võib osutuda, et ka edaspidine säilitamine, arendamine ja kasutus, sh vahepeal saadud teadustulemuste valideerimise ja täpsustamise eesmärgil on vajalik. Sellise vajaduse tekkimisel esitab uurimisgrupp vastavasisulise taotluse eetikakomiteele.

Isikuandmed pseudonüümitakse andmekogude vastutavate/ volitatud töötlejate poolt, kasutades üksnes andmeandjatele teadaolevat räsiparooli, mida nad omavahel jagavad krüpteeritult. Seda räsiparooli kasutatakse, et arvutada räsid oma andmekogudes olevatele isikukoodidele. Pseudonüümina kasutatakse saadud räsi. Andmeallikad asendavad oma andmetes isikukoodid pseudonüümidega enne andmete uuringumeeskonnale väljastamist.

Iga andmeandja väljastab valimisse kuuluvate isikute kohta oma andmebaasist pseudonüümitud andmed uuringumeeskonnale (ilma isiku otsest tuvastamist võimaldavate andmeteta). Kaudse tuvastamise riski maandamiseks rakendatakse täiendavaid andmekaitse mõjuhinnangus kirjeldatud meetmeid, sh anonüümimisrakendust vaba teksti väljadel ja otsest tuvastamist võimaldavate andmeväljade (nt aadress) välistamist. Uuringumeeskond ühendab saadud andmekomplektid pseudonüümide alusel. Nii on tagatud, et andmeandjad töötlevad vaid nende juures hoitavaid andmeid ega näe teiste asutuste andmeid.

Kirjeldatud viisil andmete vahetamisel ei liigu samaaegselt kunagi isiku otsest tuvastamist võimaldavad isikuandmed ja tema terviseandmed.

Andmete edastamiseks sõlmitakse leping iga andmekogu omaniku ja Tartu Ülikooli vahel. Andmete edastus toimub krüpteeritult läbi turvalise andmevahetusserveri, mis on üles seatud Eesti Teadusarvutuste infrastruktuuri poolt või kasutades andmeallikate tavapäraseid andmete väljastamise viise vastavalt nende sisemistele protseduurireeglitele (enamasti konkreetsele juhtivuurijale parooliga ligipääsetav kataloog andmeallika serveris, andmefail on krüpteeritud).

Analüüsiks viib uuringumeeskond need andmed eraldi virtuaalserverisse SAPU, mida haldab Eesti Teadusarvutuste

infrastruktuur.

Peame tähtsaks loodava andmestiku infoturbelist kaitset, mis nõuab kõrget ekspertiisi ja tehnilisi lahendusi, et välistada andmete hoidmine juhuslikes ebaühtlase andmekaitsega arvutites. Koostöös Eesti Teadusarvutuste infrastruktuuriga (https://etais.ee) töötab uurimisgrupp infoturbe riskide maandamiseks välja minimaalsusprintsiipi tagava andmetele ligipääsu andmise protsessi ja paigutab loodava andmestiku turvalisse andmeanalüüsi serverisüsteemi (andmepuuri) SAPU (https://sapu.cs.ut.ee). Andmete hoidmisega SAPU-s väheneb oluliselt andmelekke risk rakendades andmete eksportimisele ranged piirangud ning logides kõiki andmetega tehtavaid tegevusi kasutaja tasandil kogu andmete säilitamise ja arhiveerimise perioodil. Soovi korral tagatakse andmeandjateleandjatele ligipääs logidele. Andmete edasisel töötlemisel SAPUs ja uuringu eesmärkide täitmiseks kasutatakse OMOP mudeli jaoks loodud ja publitseeritud tööriistu (Atlas server, Usagi, Data Characterisation, PopulationLevelEstimation ja PatientLevelEstimation R-paketid, uurimisrühma arendatud tööriistad jt). Kusjuures Atlas server on OMOP andmemudelile loodud kohortide defineerimise tööriist, mis võimaldab patsiendigruppe defineerida ilma andmetele ligipääsu omamata. Seega on Atlas tööriist vajalik just minimaalsusprintsiibi tagamiseks.

SAPU serverist väljastatakse andmeid ainult vastutava uurija nõusolekul ja agregeeritud kujul nii, et on tagatud k ≥ 5 anonüümsus. SAPU serveris olevatele andmetele on ligipääs ainult käesolevas taotluses toodud uurijatel. Uurimisgrupp tagab, et andmetele ligipääs on vastavuses kehtiva seadusandlusega. Täiendavate uurimisküsimuste tekkimisel esitatakse Eesti bioeetika ja inimuuringute nõukogule ja andmeandjatele hindamiseks uus taotlus. Lisandunud uurimisküsimuste lahendamiseks antakse ligipääs andmetele lähtudes minimaalsuse printsiibist. Võimalusel viiakse lisandunud uuring läbi nii, et 1. etapi uurimisrühmale edastatakse uue uurimisküsimuse lahendamiseks kood, mille uurimisrühm läbi jooksutab ning tagastab ainult analüüsi tulemused lähtudes k ≥ 5 anonüümsusest.

Explanation of the need for de-pseudonymization [627a]

Are personal data obtained from sources other than national

No

databases? [628]

How and from which sources are personal data obtained? [628a]

How is transparency in personal data processing ensured? [629]

Kogu analüüs toimub turvalises SAPU serveris. Kogu analüüs toimub turvalises SAPU serveris. Tegu on
virtuaalse serveriga, millele ligipääs toimub üle veebi virtuaalse Linux töölaua. Lahenduse eripäraks on, et kõik serveris toimuvad operatsioonid on logitud ning pole võimalik andmeid välja
kopeerida või tõsta ilma vastava loata.

Uuritavatega ühendust ei võeta, sest uurijad ei tea, kes on 30% juhuvalimi seas.

6.3. Overview of personal data processing

Which personal data are processed in the study? [631]

Uuringus töödeldakse olemasolevaid pseudonüümitud terviseandmeid ajavahemikust 2012–2026.

Tervise infosüsteemi andmed: diagnoosid, raviskeemid, saatekirjad ja saatekirja vastused, laborianalüüside tulemused, immuniseerimise teatised ning vabatekstina patsiendi kaebused, üldseisund, allergiad ja ravi kõrvalmõjud.

Tervisekassa andmekogu andmed: raviarved, kindlustuskaitse, tervishoiuteenused ja teenuste hinnad.

Retseptikeskuse andmed: välja kirjutatud ravimid ja ravimite väljaostmine.

Vähiregistri andmed.

Surma põhjuste registri andmed: surma kuupäev ja spetsialisti kinnitatud surma põhjus.

Andmed soo, vanuse ja erinevate haiguste kohta.

Tegemist on eriliiki isikuandmetega. Andmeallikad asendavad isikukoodid pseudonüümidega enne andmete uuringumeeskonnale väljastamist.

Does the study involve the processing of health data or other

Yes

special categories of personal data? [632]

Explanation of special categories of personal data [632a]

Tervise infosüsteemi epikriisides on kirjas diagnoosid, raviskeemid, saatekirjades ja saatekirja vastustes ka laborianalüüside tulemused, kuid info võib olla puudulik. Epikriisides sisaldub vaid haigusloo kokkuvõte, mitte aga kõik teostatud uuringud või väljakirjutatud ravimid, mis mitmetes uuringutes on oluline info. Teiselt poolt on epikriisid ainsad tervisedokumendid, kus sisaldub vabas vormis kirja pandud info patsiendi kaebuste, üldseisundi, allergiate, ravi kõrvalmõjude jms kohta. Tekstilistest osadest suudame eraldada tehisintellekti meetodite abil ka muud olulist infot, mis kodeeritud väljades puudub, näiteks patsiendi kaebusi, ravimite nõrgemaid kõrvalmõjusid. Tervisekassa andmekogu info tervishoiuteenuste kohta on täielikum ja sisaldab ka teenuste hindu, kuid selle detailsus on samas madalam (mitmed tervishoiuteenused märgitud sama koodiga) ja laborianalüüside kohta puuduvad analüüsitulemused. Kõige parema pildi patsiendile välja kirjutatud ravimite osas annab retseptikeskus, lisaks on seal ka info ravimi väljaostmise kohta, mis on väga oluline indikaator ravijärgimuse hindamiseks. Samas puudub retseptikeskuses info käsimüügiravimite kohta (seda infot võib potentsiaalselt leida epikriisidest). Paljude tervisevaldkonna uuringute puhul (nt pahaloomulised kasvajad) on oluline uurida suremust, selleks on kõige kvaliteetsem info kirjas surma põhjuste registris, mis sisaldab nii surma kuupäeva kui ka spetsialisti poolt kinnitatud surma põhjuseid.

Are there more sensitive data among the personal data? [633]

Yes

Explanation of sensitive personal data [633a]

Uurimistöö käigus töödeldakse suures mahus andmesubjektide terviseandmeid ja uuritakse nende raviteekondasid, ravijärgimust ja personaliseeritud ennetusmeetmeid.
Uuringus eraldi uuritavatelt andmeid ei koguta ning andmed on retrospektiivsed.

6.4. Responsibility for personal data processing

Who is the data controller? [641]

a) The responsible institution is the sole data controller.

Explanation regarding the division of

responsibility [641a]	
Does any part of personal data processing occur without the data subject's consent? [642]	No
Justification for not using consent [644]	
Justification for using identifiable data [645a]	
Description of circumstances preventing identification [645b]	
Disclosure, sharing, storage, and reuse of personal data [646]	Andmeandjatelt saadud algandmed paigaldatakse SAPUs spetsiaalselt kirjutuskaitstud kausta, et vältida andmete juhuslikku hävimist või kahjustumist. SAPU'st tehakse regulaarselt krüpteeritud varukoopiaid Eesti Teadusarvutuste infrastruktuuri lindirobotile, mis asub füüsiliselt teises asukohas (andmekeskuses). Andmed arhiveeritakse üks aasta peale projekti lõppu. Andmete arhiveerimine üheks aastaks peale projekti lõppu on vajalik retsenseerimisel olevate artiklite publitseerimisprotsessi lõpetamiseks. Arhiveerimisperioodi lõpus, hiljemalt 31.12.2028 andmed kustutatakse. Projekti käigus võib osutuda, et ka edaspidine säilitamine, arendamine ja kasutus, sh vahepeal saadud teadustulemuste valideerimise ja täpsustamise eesmärgil on vajalik. Sellise vajaduse tekkimisel esitab uurimisgrupp vastavasisulise taotluse eetikakomiteele. Teadusuuringu tulemused publitseeritakse nii Eesti kui ka rahvusvahelistes teadusajakirjades (eelistatult vabalt kättesaadavate artiklitena), konverentsidel ja ettekannetel, üliõpilaste lõputöödes. Andmestiku metakirjeldus (kirjeldus andmeväljade kohta, ei sisalda reaalseid andmeid) OMOP-andmekogude registritesse, (nt http://portal.ehden.eu). Teadusuuringu muid tulemeid (andmestikku kirjeldus, andmehõive statistika ja -hinnangud, andmete standardiseerimise ja tekstikaeve meetodid, üleminekutabelid, teadusuuringute kokkuvõtted jms) tutvustatakse soovi korral andmeandjatele ning otsitakse koostööviise tulemite

rakendamiseks andmeandjate hallatavates andmekogudes.

Peale projekti lõppemist (dets 2027) andmed arhiveeritakse üheks aastaks, et vajadusel võimaldada arhiveerimisperioodil retsenseerimisel olevate teadusartiklite analüüsi täiendamine. Andmed hävitatakse hiljemalt 31. detsembriks 2028 kasutades hävitamise ajahetkel parimat kasutusolevat praktikat vastavalt rakendatavale infoturbestandardile (IS0/IEC 27001). Andmete hävitamine dokumenteerikse hävitamisaktis.

6.5. Risks related to personal data processing

Does the study involve systematic monitoring of individuals or collection of personal data via devices and sensors?	No
Justification for systematic monitoring [651a]	
Does the study involve profiling or automated decision-making about individuals? [652]	No
Explanation regarding profiling or automated decisions [652a]	
Are personal data from different databases, registries, or other sources combined during the study? [653]	No
Explanation regarding combining different datasets [653a]	
Does the study involve large-scale processing of sensitive or special categories of personal data? [654]	Yes
Specification of the scope of sensitive or special category data	Andmetiku aluseks on 30% juhuvalim Eestis alates 2012. aastast tervishoiuteenuseid tarbinud Eesti elanikest.

[654a]

Are personal data transferred to third countries with insufficient data protection? [655]	No
Explanation regarding countries with insufficient data protection [655a]	
Are innovative solutions or new technologies used for personal data processing in the study? [656]	No
Explanation regarding innovative solutions or technologies [656a]	
Has a data protection impact assessment been conducted for the study? [657]	Yes
Please explain why an impact assessment has not been conducted [657a]	

Additional documents

Type	File	Comment	Public
Other document		EH30 EBIN esmane otsus 1.1-12/2793	No
Data composition / Data table structure		EH30 Andmekoosseis	No
Data protection impact assessment (if applicable)		Andmekaitsealane mõjuhinnang	No
Study protocol or design		EH30 esmane taotlus	No
cover letter (additions)		Kaaskiri	No
Other document		Taotlus muudatustega kollasel taustal	No
Other document		Taotlus puhtal kujul	No

Confirmations

Role	Confirmer	Institution / Structural unit	Date of confirmation
Principal investigator	Jaak Vilo		21.07.2026 15:42:28
Validator	Jaak Vilo	Tartu Ülikool, Loodus- ja täppisteaduste valdkond, arvutiteaduse instituut	21.07.2026 15:46:23
Validator	Siret Rutiku	Tartu Ülikool	23.07.2026 08:46:36