

## Rekisterit sosiaalilääketieteellisessä tutkimuksessa

Suomessa kansallisiin tietorekistereihin on kerätty koko väestön kattavia henkilötietoja jo yli 70 vuotta. Terveiden ja hyvinvoinnin laitos, Kansaneläkelaitos, Tilastokeskus ja useat muut organisaatiot määrittelevät henkilötietojen käsittelyn perusteen ja toteutuksen rekisterinpitäjän vastuiden mukaisesti. Henkilötietoja on mahdollista hyödyntää niiden ensisijaisen käyttötarkoituksen jälkeen toissijaisesti muun muassa tieteellisessä tutkimuksessa ja innovaatiotoiminnassa sekä tiedolla johtamisessa. Eri rekistereihin tallennettuja henkilötietoja on mahdollista yhdistää henkilötunnuksen tai muun tunnuksen perusteella. Henkilötietojen toisiokäyttöä säätelee laki sosiaali- ja terveystietojen toissijaisesta käytöstä (552/2019), ja Findata tietolupaviranomaisena tai rekisterinpitäjät myöntävät tietolupia.

Rekisteritietojen toisiokäytön suurimpia vahvuuksia ovat mahdollisuus saada käyttöön koko väestön tiedot tietyltä ajanjaksolta, mikä on itse kerätyissä aineistoissa mahdotonta. Suurimmat heikkoudet liittyvät tietojen laatuun, koska rekisteritiedot eivät ole täydellisiä virheellisten ja puuttuvien tietojen vuoksi. Rekisteritietoja hyödynnetään usein havainnollistavissa tutkimuksissa, kuten kohortti- tai tapaus-verrokkitutkimuksissa sekä pitkittäistutkimuksissa, koska yksittäistä henkilöä on mahdollista seurata ajan kuluessa. Kontrolloituihin tutkimuksiin verrattaessa havainnointien tutkimusten suurimpana haasteena on se, ettei altistumista voida välttämättä määrittää tarkasti.

Euroopan Unionin (EU:n) alueella on käytössä vuonna 2018 voimaan astunut yleinen tietosuojasäätös (2016/679), joka säätelee henkilötietojen käsittelyä EU:ssa. Suomessa yksityisyyden suojaa on täydennetty kansallisella tietosuojalailla (2018/1050) ja toisiolailla. Kansallisia säädöksiä on kritisoitu siitä, että ne vaikeuttavat tieteellistä tutkimusta, koska esimerkiksi tietoja ei luovuteta konkreettisesti tutkijoille. Tämä hankaloittaa esimerkiksi pohjoismaisia yhteistutkimuksia, ja monesti joudutaan tietojen yhdistämisen sijaan tekemään meta-analyysyjä. Rekisteritutkimuksen tietolupien ja aineistojen käsittelystä aiheutuvat kustannukset ovat myös nousseet rajusti.

**Asiasanat:** rekisteritieto, rekisteritutkimus, tietosuojat, toisiokäyttö

**SARI RÄISÄNEN, KATARIINA LAINE, EEVA LIIKANEN, MIKA GISLER**

## YDINASIAIAT

- Suomessa on monia rekistereitä, joihin on kerätty sosiaalilääketieteelliseen tutkimuksen näkökulmasta merkityksellisiä henkilötietoja, joita voidaan hyödyntää ensisijaisen käyttötarkoituksen jälkeen tietoluvalla toisiokäytössä.
- Yksityisyyden suojan varmistava lainsäädäntö on muuttunut Euroopan Unionissa ja Suomessa. Tämä on muuttanut henkilötietojen tutkimuskäytön tietopyyntöjen käsittelyä ja tietolupien myöntämistä sekä aineistojen käsittelyä.

- Tulevaisuudessa tulisi pohtia miten, yksityisyyden suoja voidaan varmistaa siten, että rekisteritutkimusta voitaisiin edelleen toteuttaa sekä yksittäisissä tutkimuksissa että kansainvälisessä yhteistyössä tutkimuksen näkökulmasta joustavammin.

## JOHDANTO

Suomessa ja muissa Pohjoismaissa on ainutlaatuinen mahdollisuus hyödyntää reaali maailmasta kerättyjä koko väestön tiedot kattavia rekisteritietoja tieteelliseen tutkimukseen ja uusien innovaatioiden kehittämiseen. Suomessa yksilötasoisia henkilötietoja ja muita tietoja on kerätty kansallisiin tietorekistereihin (jäljempänä mainittaessa rekisteri) jo vuodesta 1950 lähtien. Rekisteritiedot mahdollistavat muun muassa yksilön tietojen yhdistämisen perheeseen ja biologisiin sukulaisiin, mikä vahvistaa yhteiskunnallisesti orientoituneen terveyteen, sairauteen ja hyvinvointiin kohdistuvan sosiaalilääketieteellisen tutkimuksen mahdollisuuksia. Koko väestön tiedot kattava empiirinen sosiaalilääketieteellinen tutkimus voi tuottaa merkittävää tietoa muun muassa yhteiskunnallisista muutoksista ja poliittisten päätösten vaikuttavuudesta.

Nykyään Terveyden ja hyvinvoinnin laitos (THL), Kansaneläkelaitos, Työterveyslaitos, aluehallintovirastot, Fimea ja Valvira toimivat rekisterinpitäjinä lukuisille eri tarkoituksiin perustetuille rekistereille. Rekisterinpitäjällä tarkoitetaan tässä organisaatiota, joka määrittää henkilötietojen käsittelyn tarkoituksen ja tavat. Kansaneläkelaitos ylläpitää terveydenhuollon sähköisesti kerätyn potilastiedon arkistoa (Kanta), joka kattaa kaikkien julkisten terveydenhuollon toimijoiden sekä yksityisten palveluntuottajien tarjoaman hoidon tiedot vuodesta 2014 alkaen. THL:n ylläpitämiin laaturekistereihin kerätään tietoja eri sairauksista, kuten diabeteksesta, suun ja hampaiden hoidosta sekä psykoosin hoidosta, tavoitteena parantaa hoidon laatua, tuloksia ja vaikutuksia. Suomessa toimii myös 11 biopankkia, ja Suomen Biopankkien osuuskunta (FINBB) edistää kansallisen terveys- ja biolääketieteellisen tutkimuksen kilpailukykyä muun muassa tuotta-

malla tutkimustoimintaa edistäviä palveluita jäsenilleen ja omistajilleen.

Suomessa vuonna 2019 perustettu sosiaali- ja terveysalan tietolupaviranomainen Findata (1) myöntää tietopyyntöjen perusteella tietolupia asiakas- ja rekisteritietojen toissijaiseen käyttöön toisiolain perusteella. Toisiokäyttö viittaa tietojen käyttämiseen muussa kuin niiden alkuperäisessä käyttötarkoituksessa, ja toisiolaissa määritellyt toissijaiset käyttötarkoitukset kattavat tieteellisen tutkimuksen, tilastoinnin, kehittämisen ja innovaatiotoiminnan, viranomaisohjauksen ja -valvonnan, suunnittelu- ja selvitystehtävät, opetuksen sekä tietojohtamisen. (2.)

Toisiolain tavoitteena on lisätä rekisteröityjen luottamusta henkilötietojensa käsittelyyn, jota säätelevät Euroopan Unionin tietosuojasetus (3), kansallinen tietosuojalaki (1050/2018) (4) ja laki viranomaisen toiminnan julkisuudesta (621/1999) (5). THL:n yhteydessä toimivan Findatan toimintaa ohjaa sosiaali- ja terveystieteiden ministeriö (STM), joka nimittää yksikön johtajan ja ohjausryhmän (1).

Tämän artikkelin tarkoituksena on kuvata rekisteritutkimuksen menetelmiä ja mahdollisuuksia Suomessa toisiolain voimaantulon jälkeen. Tavoitteena on edistää rekisteritietojen toisiokäyttöä tieteellisessä tutkimuksessa ja innovaatiotoiminnassa.

## SOSIAALI- JA TERVEYDENHUOLLON REKISTERINPITÄJÄT JA TIETOREKISTERIT

Taulukossa 1 kuvataan Terveyden ja hyvinvoinnin laitoksen (THL), Työterveyslaitoksen, Tilastokeskuksen, Lääkealan turvallisuus- ja kehittämiskeskus Fimean, Kansaneläkelaitos Kelan, aluehallintovirastojen ja Valviran rekistereitä, jotka sisältävät koko väestöä tai tiettyä väestöryhmää koskevia yksikkötasoisia tietoja, joita päivitetään.

Taulukko 1. Sosiaali- ja terveydenhuollon aihealueen keskeiset rekisterinpitäjät ja rekisterit sekä niiden tietosisältö ja tiedonkeruu Suomessa

Rekisterin pitäjä	Rekisterit	Tietosisältö	Tiedonkeruu
Terveyden ja hyvinvoinnin laitos ja Syöpäyhdistys	Syöpä- ja joukkotarkastusrekisteri	Suomessa todetut syöpätapaukset seulonta-asetuksen mukainen syövän joukkotarkastustoiminta	1953/1968 kohdunkaulasyövän seulonta 1987 rintasyövän seulonta 2022 suolistosyövän seulonta
THL	Epämuodostumarrekisteri	Sikiöaikana ja syntymän jälkeen todetut epämuodostumat	1963
THL	Terveydenhuollon hoitoilmoitus, rekisteri, Terveys Hilmo (vuoteen 1993 asti nimeltään poistoilmoitusrekisteri)	Sairaaloiden ja terveyskeskusten vuodeosastohoidon sekä julkisen erikoissairaanhoidon avohoidon toiminnasta ja niissä hoidossa olleista potilaista sekä kotihoidosta	1967/1994 päiväkirurgia 1998 sairaaloiden poliklinikkakäynnit
THL	Implanttirekisteri	Tiedot ihmiseen asetettavista nivelproteeseista ja tiedon hammasimplanteista	1990/1994 hammasimplanttien keruu lopetettu 2016
THL / Näkövammaisten liitto ry	Näkövamma-rekisteri	Tiedot näkövammasta ja sen syistä	1983
THL	Raskauden-keskeyttämisrekisteri	Tiedot laillisesti suoritetuista raskauden keskeytyksistä	1950, rekisteritietoja saatavissa vuodesta 1983 lähtien
THL	Steriloimisrekisteri	Tiedot laillisesti suoritetuista steriloinneista	1939, rekisteritietoja saatavissa vuodesta 1987 lähtien
THL	Syntyneiden lasten rekisteri	Tiedot kaikista Suomessa synnyttäneistä äideistä ja elävinä ja kuolleina syntyneistä lapsista	1987
THL	Lastensuojelurekisteri	Tiedot kaikista suomalaisista kodin ulkopuolelle sijoitetuista lapsista ja nuorista	1991
THL	Sosiaalihuollon hoitoilmoitusrekisteri, sosiaali-Hilmo	Tietoa sosiaalihuollon laitoshoidosta ja ympärivuorokautisesta sekä ei-ympäri-vuorokautisesta asumispalvelusta, laitosten toiminnasta ja niiden asukkaista	1994
THL	Tartuntatauti-rekisteri	Laboratorioiden ja lääkäreiden tartuntatauti-ilmoitukset	1995
THL	Toimeentulotukirekisteri	Tiedot kaikista kotitalouksista (perheistä ja yksinasuvista avunsaajista), jotka ovat saaneet toimeentulotukea toimeentulotukilain mukaan	1997
THL	Perusterveydenhuollon avohoidon hoitoilmoitusrekisteri Avohilmo	Tiedot perusterveydenhuollon avohoidosta, kotihoidosta ja työterveyshuollosta	2011
THL	Laaturekisterit: diabetes, HIV, munuaistaudit, psykoosi, selkä, suun ja hampaiden hoito, sydäntehoahoito, tulehdukselliset reumasairaudet	Tietoa asiakkaiden tai potilaiden saaman hoidon laadusta, tuloksista ja vaikutuksista	2023

THL	Sosiaalihuollon seurantarekisteri	Tietoja, jotka ovat tärkeitä sosiaalipalvelujen kehittämisen ja seurannan kannalta. Näihin tietoihin kuuluvat asiakkaan ja järjestäjän yksilölliset asiakkuusjaksot palvelutehtävissä, lastensuojelun asiakkuusjaksot, palvelupäätökset sekä muut tiedot palvelun saamisesta.	2023
THL	Ensihoitopalvelun rekisteri	Kanta-palveluiden potilastiedot, ensihoitopalvelun toiminnan tiedot turvallisuusviranomaisten kenttäjohtojärjestelmästä sekä hätäilmoitusten ja tehtävätiedot hätäkeskustietojärjestelmästä	2023–2024
THL	Hedelmöityshoitorekisteri	Klinikoiden syklikohtaiset tiedot annetuista hoidoista sekä seurantatiedot hoidosta alkaneista raskauksista ja syntyneistä lapsista	2026
Työterveyslaitos	Biologisten altistumismittausten rekisteri	Altistujatiedot ja tiedot, altistumisen laadusta, kestosta ja mittauksen tilaajasta	1950
Työterveyslaitos	Työperäisten sairauksien rekisteri, TPSR	Tiedot työperäisten sairauksien laadusta, vakavuudesta ja seurauksista sekä vakuutus päätöksistä	1964
Työterveyslaitos	ASA-rekisteri	Tiedot työssä syöpäsairauden vaaraa aiheuttaville aineille ja menetelmille sekä perimää vaurioitaville aineille altistumisesta	1979
Tilastokeskus	Kuolemansyyrekisteri	Tiedot kuolemista ja kuolemansyistä tautiluokituksen tunnuksina ja tekstinä	1936, rekisteritietoja saatavilla vuodesta 1969 lähtien
Fimea	Rokotusten haittavaikutusrekisteri	Tiedot rokotteiden haittavaikutuksista	1986
Fimea	Lääkkeiden haittavaikutusrekisteri	Tiedot lääkkeiden haittavaikutuksista	1966
Fimea	Erityisluparekisteri	Tiedot lääkeaineiden erityislupahakemuksista	2011
Fimea	Vaaratilanne rekisteri	Tietoa lääkinnällisten laitteiden aiheuttamista vaaratilanteista	1995
Kela	Etuusrekisteri	Tietoa etuuden saajista ja etuuksista sekä opiskelijaterveydenhuollon käytöstä	1964
Kela	Reseptitietokanta	Tiedot lääkemääräyksistä ja lääkkeiden luovutuksista	Julkisen terveydenhuollon tiedot 3/2013 alkaen ja yksityinen terveydenhuolto 3/2014 alkaen
Kela	Kanta potilastiedon arkisto	Terveydenhuollossa sähköisesti kerätyt potilastiedot	Julkiset palvelun tuottajat 2014 alkaen ja yksityiset viimeistään 11/2021 alkaen
Aluehallintovirastot		Tiedot sosiaali- ja terveydenhuollon toimintayksiköiden ja ammattihenkilöiden valvonnasta ja ohjauksesta	2011
Valviva	Terhi Suosikki	Tietoa sosiaali- tai terveydenhuollon ammattihenkilön ammattipätevyydestä	Terhikki 1994 Suosikki 2016

Terveyden ja hyvinvoinnin laitos on rekisterinpitäjänä monissa sosiaali- ja terveydenhuollon rekistereissä, joista vanhin on 1953 perustettu Syöpärekisteri, jonka tietoja ylläpitää ja käsittelee Suomen Syöpäyhdistys. Syöpärekisteriin tallennetaan tietoja kaikista syöpätapauksista, kuten tiedot syöpäkasvaimesta, potilaan saamasta hoidosta ja sairastuneen taustatietoja. Syöpärekisteriin tallennetaan myös tiedot kohdunkaulan syövän sekä rinta- ja suolistosyövän seulonnoista. Suolistosyövän seulonnat aloitettiin Suomessa 2022.

Hilmo on valtakunnallinen sosiaali- ja terveydenhuollontiedonkeruu- ja raportointijärjestelmä, johon THL kerää sosiaali- ja terveydenhuollon potilas- ja asiakastietojärjestelmistä tiedot sairaaloiden ja terveyskeskusten vuodeosastohoidosta sekä julkisen erikoissairaanhoidon avohoidon toiminnasta ja potilaista. Tiedot tallennetaan sosiaali-Hilmoon, terveys-Hilmoon ja perusterveydenhuollon kattavaan Avohilmoon. THL:n julkaisemissa oppaissa kuvataan edellä kuvattujen rekistereiden tiedonkeruun ja tietojen lähettämisen muutoksista.

Raskaudenkeskeytysrekisteriin on kerätty tiedot kaikista laillisesti suoritetuista raskaudenkeskeytyksistä, keskeytyksen perusteista ja päätöksenteosta vuodesta 1983 alkaen. Laki raskaudenkeskeyttämisestä (1097/2022) muuttui syyskuussa 2023, kun raskaus voidaan keskeyttää kahdennentoista raskausviikon loppuun saakka raskaana olevan pyynnöstä. Tämä aiheutti muutoksia Terveyden ja hyvinvoinnin laitoksen tiedonkeruuseen ja rekisterin tietosisältöön. Hedelmöityshoitorekisterin perustaminen käynnistyi vuonna 2024, ja tiedonkeruu alkaa vuonna 2026.

Vuonna 1987 perustettuun syntyneiden lasten rekisteriin kerätään tiedot kaikista Suomessa elävänä tai kuolleena syntyneistä lapsista, joiden syntymäpaino on vähintään 500 grammaa tai kun raskauden kesto on vähintään 22 viikkoa sekä tiedot synnyttäjistä. Syntyneiden lasten rekisterin tietosisältöjä kehitetään yhdessä synnytysairaaloiden ja alan asiantuntijoiden kanssa. Tietoja hyödynnetään laajasti seksuaali- ja lisääntymisterveyden tutkimuksessa ja kehittämisessä. Kaikissa Pohjoismaissa on vastaavat rekisterit, jotka tekevät paljon yhteistyötä.

Suomessa on tällä hetkellä yhdeksän laaturrekisteriä (Taulukko 1), joihin terveydenhuollon ammattilaiset kirjaavat tietoja annetusta hoi-

dosta, hoidon tuloksista ja vaikutuksista. Tällä hetkellä Terveyden ja hyvinvoinnin laitos ylläpitää diabetes-, HIV-, munuaistauti-, selkä- ja sydänrekistereitä ja näiden lisäksi laaturekistereitä on myös psykoosin, suun ja hampaiden, tehohoidon ja tulehdussellisten reumasairauksien hoidon laadun kehittämistä varten. Laaturekistereiden tietosisältöä kehitetään edelleen ja tulevaisuudessa tietoa hoidon vaikuttavuudesta tullaan keräämään myös suoraan potilailta. Eturauhassyövän hoidon laaturekisterin perustamista selvitetään.

Tilastokeskuksen kuolemansyytiedot kuuluvat toisiolain piiriin, joten lupa ja aineisto haetaan Findatalta, mikäli tutkimus kattaa vähintään kahden rekisterinpitäjän tietoja, Kanta-palveluiden tietoja tai yksityisen sosiaali- tai terveydenhuollon palvelunjärjestäjän tietoja. Muut Tilastokeskuksen hallinnoimat tiedot, kuten sosioekonomista asemaa kuvaavat koulutus- ja ansiotulotiedot sekä tiedot perheestä ja suvusta haetaan suoraan Tilastokeskukselta.

Kansaneläkelaitos Kelan etuusrekisteriin kerätään eri viranomaisilta ja muilta toimijoilta, kuten vakuutuslaitoksilta tietoja etuuksista ja niiden saajista sekä etuuksien käsittely- ja tilastotietoja. Lisäksi etuusrekisteri sisältää tietoja opiskeluterveydenhuoltoa käyttävistä opiskelijoista ja asiakaspalvelusta. Kelan ylläpitämään sähköiseen Kanta-potilastiedon arkistoon tallennetaan terveydenhuollon palveluissa sähköisesti kerätyt potilastiedot, joita voidaan luovuttaa terveydenhuollon rekisterinpitäjien käyttöön rekisteröidyn luvalla. Kantapalveluiden käyttöönotto edellyttää sosiaali- ja terveydenhuollon organisaatiolta rakenteisen kirjaamisen käyttöönottoa potilastiedon kirjaamisessa.

## BIOPANKKIAINEISTOT

Suomessa on tällä hetkellä 11 biopankkia, joiden toimintaa ohjaa ja valvoo Fimea. Biopankit toimivat hyvinvointialueiden, yliopistojen, Terveyden ja hyvinvoinnin laitoksen, Veripalvelun ja Terveystalon yhteydessä ja keräävät suostumukseen perustuen näytteitä alueellisesti tai valtakunnallisesti. Biopankkien yhteistyötä ja terveys- ja biolääketieteellisen tutkimuksen kilpailukykyä kehittää Biopankkien Osuuskunta Suomi (FINBB). FINBB:n digitaalisen Fingenious®-palvelun kautta biopankkeihin tallennettuihin näytteisiin sekä fenotyyppi- ja genotyyppitietoihin voi hakea käyttöilupaa

korkeatasoisiin lääketieteellisiin tutkimuksiin tai tuotekehitysprojekteihin, joiden tavoitteena on edistää väestön terveyttä ja sairauksien hoitoa. Fingenious® toimii Findatan tapaan yhden luukun periaatteella, ja biopankkinäytteet ovat tutkijoiden käytettävissä maailmanlaajuisesti. Biopankkinäytteisiin voi yhdistää rekisteritietoja Findatan kautta tutkimussuunnitelman mukaisesti.

Esimerkiksi FinnGen-tutkimushanke on laaja genomitutkimus, jossa on tavoitteena analysoida yli 500 000 suomalaisen biopankkinäytteitä yhdistäen geneettistä ja terveystietoa sairauksien mekanismien ja alttiuksien ymmärtämiseksi (6).

## FINDATAN TOIMINTA

Findata käsittelee toisiolain mukaisesti tietopyyntöjä ja myöntää siinä määriteltyihin tarkoituksiin tietolupia sekä toteuttaa aineistojen muodostamisen ja tietojen yhdistämisen, kun tietoja halutaan useammalta kuin yhdeltä rekisterinpitäjältä. Jos aineisto muodostuu vain yhden rekisterinpitäjän tiedoista, sen voi muodostaa myös yksittäinen rekisterinpitäjä (1).

Osa rekisterinpitäjistä, kuten THL on siirtänyt lupahallintansa kokonaan Findatalle muiden kuin THL Biopankin näytteiden ja tietojen osalta. Myös aluehallintovirastojen ja Sosiaali- ja terveysalan lupa- ja valvontavirasto Valviran lupahallinnon hoitaa Findata. Näiden rekisterinpitäjien aineistoja koskevat tietolupapyyntöt haetaan aina Findatalta, vaikka tietoja ei yhdistetä toisen rekisterinpitäjän aineistoihin. Julkista rekisterinpitäjistä Digi- ja väestötietovirasto, Eläketurvakeskus ja Tilastokeskus eivät voi siirtää lupahallintaa Findatalle. (1.)

Findatan sivustolla kuvataan tietopyyntöjen toteutusta erilaisissa käyttötapauksissa ja eri rekisterinpitäjien kohdalla (1). Tietopyyntöön tarvitaan aina tutkimus- tai hankesuunnitelma, joka sisältää kuvauksen henkilötietojen käsittelyn tarkoituksesta, rekisterinpitäjistä, henkilötietojen käsittelijöistä sekä tietosuojan ja tietoturvan varmistamisesta. Lisäksi tarvitaan kuvaus pyydettävistä tiedoista ja aineistonhallinnasta pyydettyjen tietojen koko elinkaaren ajalta. (1,2,4.) Kun Findatalta haettavia tietoja on tarkoitus yhdistää tutkittavan suostumuksella kerättävään aineistoon, on hakemuksen liitteenä toimitettava tutkittavan tiedote- ja suostumusmallit (1,2).

Yksilötasoista tietoa sisältävät aineistot ovat henkilötietoa ja niiden käsittely on mahdollista

vain Findatan tietoturvalisessä Kapseli-käyttöympäristössä tai muussa tietoturvalisessä ympäristössä, joista Valvira ylläpitää julkista rekisteriä (2). Tietosuojan varmistaminen koskee myös tulosten julkaisemista, ja Findatan sivustolla on anonymiteetin varmistamista koskevia ohjeita ja lisäksi se tarjoaa maksutonta palvelua tulosten anonymiteetin varmistamiseen (1). Tilastoaineistoja, joista yksilöä ei voida tunnistaa, voi käsitellä tarvittaessa myös muussa käyttöympäristössä.

## REKISTERIAINEISTOJEN KUVAILUTIEDOT

Findata tuottaa yhdessä edellä kuvattujen rekisterinpitäjien kanssa tietoa aineistojen kuvailu- eli metatietoja, joiden perusteella on mahdollista arvioida tietojen hyödyntämisen mahdollisuuksia toisiolain mukaisiin käyttötarkoituksiin. Kuvailutiedot on koottu aineistokatalogiin, josta niitä on mahdollista hakea rekisterinpitäjän, aineiston tai muuttujan perusteella. Rekistereiden tietosisällöstä ja tietojen keruusta saa lisätietoa rekisterinpitäjän laatimista tietosuojailmoituksista. (1.)

## REKISTERITIE TOJEN LAATU

Rekisteritiedon laatu on suurelta osin riippuvainen siitä, miten tiedot on alun perin tallennettu. Jos asiakas- ja potilastiedot kirjataan virheellisesti tai puutteellisesti niiden ensisijaista käyttötarkoitusta varten, niiden laatu on heikkoa myös toisijaisessa käyttötarkoituksessa (7,8). Yhteisesti sovitut tavat järjestää, esittää ja käyttää tietoa ovat avainasemassa tiedon laadun varmistamisessa sosiaali- ja terveydenhuollossa. Tiedon laadun varmistamisessa on tärkeää säilyttää tiedon semanttinen merkitys eli ymmärrettävyys ja tulkin sekä varmistaa tiedon sujuva siirtäminen tietojärjestelmien välillä. Hyvä yhteentoimivuus parantaa tiedon laatua vähentämällä virheitä ja epäjohtomukaisuuksia tiedon siirrossa ja käytössä, lisäämällä tiedon vaikuttavuutta ja tuomalla säästöjä. (9.)

Rekisteritietojen laatu voi vaihdella eri rekistereiden välillä ja jopa yksittäisen rekisterin sisällä eri muuttujien osalta (10,11). Yksittäisen tutkimuksen luotettavuus paranee, kun tutkijoilla on arviointitietoa käyttämänsä tietolähteen laadusta. Tietojen laatua voidaan arvioida vertaamalla tietorekisterin tietosisältöä alkuperäiseen sosiaali- ja terveydenhuollon kirjaamiseen potilasdokumentissa tai keräämällä tietoa suoraan asiakkailta tai potilailta (11,12). Viimeisen kah-

denkymmenen vuoden aikana on julkaistu useita tutkimuksia, jotka ovat keskittyneet arvioimaan yksittäisen diagnoosin tai toimenpiteen rekisteröinnin laatua tiettyssä rekisterissä (10,13–15). Rekisterinpitäjät eivät säännöllisesti tee kattavuus- tai laatu-tutkimuksia, vaikka julkaisevatkin eri tiedonkeruistaan laatuselosteita, joissa arvioidaan tietojen luotettavuutta ja sopivuutta eri käyttötarkoituksiin Suomen virallisen tilaston (SVT) laatuselostesuosituksen mukaan.

## REKISTERIAINEISTON TUTKIMUKÄYTTÖÖN LIITTYVÄT VAHVUUDET JA HEIKKOUEDET

Koko väestön tiedot sisältävien rekisteritietojen tutkimuskäytöllä on useita vahvuuksia, koska suuremmalla otoskoolla voidaan vähentää sattumanvaraista virhettä sekä parantaa tulosten tarkkuutta ja luotettavuutta. Aineiston pieni koko voi johtaa otosharhaan etenkin, jos se ei edusta luotettavasti koko populaatiota, mikä voi johtaa virheellisiin johtopäätöksiin ja vääristää tutkimustuloksia. Lisäksi pieni otoskoko voi johtaa tilanteeseen, jossa tutkimuksella ei ole riittävää voimaa havaita tilastollisesti merkitsevää eroa, jolloin todellinen yhteys tai vaikutus saattaa jäädä havaitsematta. (8,16.)

Rekisteritiedot mahdollistavat myös pitkäaikaisseurannan, jolloin voidaan tutkia esimerkiksi syöpäseulontojen vaikuttavuutta, sairauksien ilmaantuvuutta ja esiintyvyyttä väestötasolla sekä seurata tutkimukseen valittua joukkoa (kohorttia). Rekisteriaineiston merkittävänä vahvuutena on myös se, että tiedot on kerätty valmiiksi, mikä vähentää tutkimuksen kustannuksia ja aineiston keräämiseen tarvittavia resursseja. Eri rekistereiden tietoja voidaan yhdistää henkilötunnuksen tai muun mahdollisen tunnuksen kuten organisaatiotunnuksen avulla. Lisäksi rekisteritutkimuksella on tärkeä rooli sosiaali- ja terveydenhuollon palvelujärjestelmän toiminnan tutkimuksessa. (7,8,11.)

Rekisteritietojen tutkimuskäyttöön liittyy myös heikkouksia. Rekisteritiedot on kerätty lähes aina alun perin muuhun kuin tutkimukseen, joten tiedot eivät välttämättä ole täydellisiä niiden toissijaisessa käytössä. Rekisteritiedoissa käytetyt muuttujien määrittelyt voivat vaihdella eri rekistereiden ja ajan myötä esimerkiksi yhteiskunnallisten muutosten vuoksi, mikä voi vaikeuttaa tietojen vertailua ja yhdistämistä. Tästä esimerkkinä mahdollisuus korjata täysi-ikäisenä

väestötietojärjestelmään tieto omasta sukupuolesta siten, että se vastaa omaa sukupuoli-identiteettiä. Kaikkia tarvittavia tietoja ei välttämättä ole saatavilla rekistereistä, tai tiedot voivat olla puutteellisia tai epätarkkoja. Rekisteritiedot voivat sisältää vääristymiä, kuten valikoitumisvääristymää, jossa tiettyjen ryhmien tiedot ovat ylitai aliedustettuina. (7,17.)

Suuren otoskoon tutkimuksissa p-arvon tulokintaan liittyy haasteita, sillä suuressa aineistossa havaitaan useammin tilastollisesti merkitseviä eroja, jotka eivät kuitenkaan välttämättä ole kliinisesti tai kansanterveydellisesti merkityksellisiä. P-arvon sijaan luottamusvälin raportoiminen voi olla luotettavampi menetelmä tulosten kliinisen tai todellisen merkitsevyyden arvioimiseksi. (18.)

Varsinkin suurten aineistojen (big data) tai koneoppimishankkeissa (machine learning) tietojen minimointiperiaate aiheuttaa ongelmia. Tutkimusryhmä ei välttämättä saa käyttöönsä kaikkia tarvittavia tietoja, esimerkiksi rekisterin kaikkia muuttujia tai tutkimusryhmään kuuluvien kaikkia ICD-10-koodeja tai lääkemääräyksiä anatomic-terapeuttis-kemiallisine (ATC) -koodeineen. Tietolupia käsittelevillä ja tutkimusta suunnittelevilla voi olla eri käsitys siitä, ovatko henkilötiedot asianmukaisia, olennaisia ja rajoitettuja siihen, mikä on tarpeellista suhteessa niihin tarkoituksiin, joita varten niitä käsitellään. Henkilötietojen käyttö tutkimuksessa edellyttää tietosuojaan huomioimista ja asianmukaisten lupien hankkimista, mikä voi olla aikaa vievää ja monimutkaista (3,4). Nykyisin rekisteritietojen tutkimuskäytön luvat ja aineistojen muodostaminen ja tietoturvallisten analyysiympäristöjen käyttö voi olla kallista. Henkilötietoja ei voi siirtää toiseen maahan, mikä hankaloittaa kansainvälistä tutkimusyhteistyötä. (2.)

## HAVAINNOIVAT TUTKIMUSASETELMAT SOSIAALILÄÄKETIETEELLISESSÄ REKISTERITUTKIMUKSISSA

Sosiaalilääketieteellisissä tutkimuksissa rekisteritietoja hyödynnetään erityisesti havainnoivissa tutkimusasetelmissa, joita ovat poikkileikkaus-, ekologinen ja kohorttitutkimus, sekä tapaus-verrokkitutkimus. Poikkileikkaustutkimuksessa tarkastellaan kiinnostuksen kohteena olevan tekijän esiintymistä tai ilmaantuvuutta väestössä tiettyä ajankohtana (11), kun taas ekologisessa tutkimusasetelmassa tarkastellaan tekijän ilmaantuvuudessa tai esiintyvyydessä tapahtunutta muutosta

tietyllä alueella. Poikkileikkaustutkimuksesta esimerkkinä on tanskalainen tutkimus sosioekonomisista terveyseroista ja suomalainen tutkimus sosioekonomisen aseman yhteydestä terveyspalveluiden käyttöön (19). Kohorttitutkimuksessa seurataan tarkoin määriteltyä ryhmää ajan kuluessa (20,21), ja tapaus-verrokkitutkimuksessa verrataan esimerkiksi puolison syöpädiagnoosin vaikutuksia terveen puolison työmarkkinoille osallistumiseen eri ajanjaksoina Suomessa (22).

## HAVAINNOLLISTAVIEN TUTKIMUSASETELMIEN VAHVUUDET JA HEIKKOUEDET

Havainnollistavien tutkimusasetelmien vahvuutena on mahdollisuus testata tosielämän havainnoilla satunnaistettujen kontrolloitujen tutkimusten (randomized controlled trials, RCT) tuloksia, joiden tuottamaa tietoa pidetään ”kultaisena standardina”. Havainnollistavissa tutkimusasetelmissä oleellisena heikkoutena on altistumistiedon, kuten tupakoinnin, ruokavalion tai liikunnan harastamisen määrittäminen tutkittavilta itseltään, mikä on merkittävä heikkous RCT-tutkimuksiin verrattaessa. Lisäksi elintavoissa kuten tupakoinnissa voi tapahtua muutoksia, joita ei välttämättä ole rekisteröity luotettavasti. Kun päätetapahtumana on vakava sairaus, on ilmeistä, että tieto sairastumisesta vaikuttaa altistumisen raportointiin, koska esimerkiksi syöpään sairastunut henkilö voi miettiä altistumistaan toisin kuin sairastumaton henkilö. Tämä on erityinen heikkous varsinkin tapaus-verrokkitutkimuksessa, jossa altistumistiedot kerätään vasta päätetapahtuman jälkeen. Havainnollistavat tutkimusasetelmat voivat kuitenkin täydentää RCT-tutkimuksia ja tarjota vahvaa näyttöä ja laajemman näkökulman sekä paremman ymmärryksen reaali maailman ilmiöistä. (7,17.)

Myös RCT-tutkimuksiin voi liittyä heikkouksia verrattaessa havainnollistaviin tutkimusasetelmiin. Ensinnäkin RCT-tutkimukset voivat olla kalliita ja aikaa vieviä toteuttaa, mikä rajoittaa niiden käyttöä laajoissa populaatioissa ja kun seuranta-aika on pitkä. Toiseksi RCT-tutkimukset saattavat olla liian kontrolloituja ja keino-tekoisia, jolloin niiden tulokset eivät välttämättä ole yleistettävissä reaali maailman tilanteisiin. Kolmanneksi RCT-tutkimuksiin voi liittyä eettisiä haasteita, erityisesti silloin, kun tutkittavien ryhmien välillä on merkittäviä eroja hoidon saatavuudessa tai laadussa. Lisäksi RCT-tutkimuk-

set eivät aina pysty ottamaan huomioon kaikkia mahdollisia muuttujia ja kontekstuaalisia tekijöitä, jotka voivat vaikuttaa tutkimustuloksiin. (23.)

## LOPUKSI

Yhteenvetona rekisteritutkimuksen nykytilasta voidaan todeta, että toisiolain soveltamisella on ollut merkittäviä vaikutuksia rekisteritutkimukseen Suomessa. Toisiolain perusteella henkilötietojen toissijaiselle käsittelylle tieteellisessä tutkimuksessa ja innovaatio toiminnassa tarvitaan joko yksittäisen rekisterinpitäjän tai Findatan myöntämä tietolupa, ja henkilötietoja voi käsitellä vain erikseen tätä käyttöä varten auditoiduissa tietoturvallisissa käyttöympäristöissä. (1,2.) Tällä muutoksella on selkeytetty ja yhtenäistetty henkilötietojen tutkimuslupakäytäntöjä, mutta toisaalta lupien saaminen on hidastunut ja tutkimuksesta on tullut tutkijoille huomattavan kallista, koska tietoluvat, aineistojen muodostaminen ja anonymisointi eli henkilötietojen karkeistaminen niin, ettei henkilöä voida tunnistaa niistä sekä tietoturvalliset käyttöympäristöt ovat lisänneet huomattavasti rekisteritutkimukseen liittyviä kustannuksia. Tämä koskee erityisesti kliinistä rekisteritutkimusta, jota usein tehdään ilman erillistä rahoitusta työajan jälkeen. Haittaa aiheutuu myös siitä, että pseudonymisointijakaan eli ilman lisätietoja tunnistamattomaksi karkeistettuja tietoja tai anonyymejä henkilötietoja ei luovuteta ulkomaille, mikä tekee mahdottomaksi esimerkiksi pohjoismaisen yhteistyön. Pohjoismainen yhteistyö rekistereiden tietosisältöjen kehittämisessä ja tutkimuksessa on tärkeää sosiaali- ja terveydenhuollon palveluiden laadun arvioimisessa ja kehittämisessä, mikä vaikuttaa suoraan terveydenhuollon vaikuttavuuteen ja kustannuksiin. (24.)

Toisiolain tulkinta henkilötietojen käsittelyssä on osoittautunut kokonaisuutena monimutkaiseksi, koska esimerkiksi henkilön itsensä antama suostumus ei välttämättä mahdollista henkilötietojen toisiokäyttöä. Suomen tietosuoja laki on Euroopan unionin tietosuoja-asetukseen verrattuna tiukempi, mikä heijastuu nykyisiin toisiolain mukaisiin rekisteritutkimuksen käytäntöihin, joita tässä artikkelissa on kuvattu. (24.)

Pääministeri Orpon hallitusohjelmaan on kirjattu tavoitteeksi tutkimuslainsäädännön kokonaisuuden selkeyttäminen ja yhtenäistäminen.

Toisiolain osalta on keskusteltu tarpeesta lisätä joustavuutta ja ketteryyttä rekisteritietojen lupakäytäntöihin ja tietoturvallisten käyttöympäristöjen perustamiseen sekä lisätä tutkimusaineistojen muodostamisen laskuttamisesta. Odotettavissa on siis muutoksia viimeistään vuoden 2027 aikana.

Euroopan unionissa saatiin keväällä 2024 aikaan päätös eurooppalaisen terveysdata-avaruuden (EHDS) tavoitteista. Eurooppalaisen terveysdata-avaruuden tavoitteena on antaa yksilöille mahdollisuus hallita omia terveystietojaan ja helpottaa terveyspalveluiden tarjoamista tietoturvalisella terveystietojen käsittelyllä. Tavoitteena on myös luoda tehokas ja turvallinen digitaalinen järjestelmä terveystietojen toisiokäytölle tutkimukselle ja uusille innovaatioille, jotka edistävät väestön terveyttä ja hyvinvointia sekä terveydenhuollon palveluiden vaikuttavuuden ja laadun kehittämistä. Euroopan unionin jäsenmaiden kansalliset tietosuojasäädökset voivat kuitenkin heikentää terveysdata-avaruuden käyttöönottoa ja siltä odotettuja innovaatioita ja hyötyjä. (25.)

## LÄHTEET

- (1) Findata. Luettu 10.9.2024. <https://findata.fi/>
- (2) Laki sosiaali- ja terveystietojen toissijaisesta käytöstä 552/2019. Luettu 3.9.2024. <https://www.finlex.fi/fi/laki/alkup/2019/20190552>.
- (3) Yleinen tietosuoja-asetus (GDPR) 2018/1725. Luettu 3.9.2024. <https://eur-lex.europa.eu/FI/legal-content/summary/general-data-protection-regulation-gdpr.html>
- (4) Tietosuoja laki 5.12.2018/1050. Luettu 4.9.2024. <https://www.finlex.fi/fi/laki/ajantasa/2018/20181050>
- (5) Laki viranomaisten toiminnan julkisuudesta 21.5.1999/621. Luettu 4.9.2024. <https://www.finlex.fi/fi/laki/ajantasa/1999/19990621>
- (6) Kurki MI, Karjalainen J, Palta P, ym. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 2023;613:508–518. doi: 10.1038/s41586-022-05473-8
- (7) Vuorela P, Malmivaara A, Komulainen J, ym. Miten arvioin ja hyödynnän havainnoivan tutkimuksen tuottamaa tietoa? *Duodecim* 2014;130:1545–1550.
- (8) Peltola M, Juntunen M, Häkkinen U, ym. A methodological approach for register-based evaluation of cost and outcomes in health care. *Ann Med* 2011;43:S4–13. doi: 10.3109/07853890.2011.586364
- (9) Valtiovarainministeriö. Suositus tiedonhallintamallista. Luettu 5.9.2024. <https://urn.fi/URN:ISBN:978-952-367-445-5>
- (10) Toivonen E, Taurio K, Kortelainen E, ym. Validation of the Finnish Care register for Health Care diagnoses for preeclampsia, gestational diabetes and preterm delivery. *Pregnancy Hypertens* 2024;35:26–9. doi: 10.1016/j.preghy.2023.12.001
- (11) Kesmodel US. Cross-sectional studies – what are they good for? *Acta Obstet Gynecol Scand* 2018;97:388–393. doi: 10.1111/aogs.13331
- (12) Helle N, Niinimäki M, Linnakaari R, ym. National register data are of value in studies on miscarriage – Validation of the healthcare register data in Finland. *Acta Obstet Gynecol Scand* 2022;101:1245–52. <https://doi.org/10.1111/aogs.14445>
- (13) Al-Zirqi I, Stray-Pedersen B, Forsén L, ym. The NUR group. Validation study of uterine rupture registration in the Medical Birth Registry of Norway. *Acta Obstet Gynecol Scand* 2013;92:1086–93. doi: 10.1111/aogs.12148
- (14) Baghestan E, Børdahl PE, Rasmussen SA, ym. A validation of the diagnosis of obstetric sphincter tears in two Norwegian databases, the Medical Birth Registry and the Patient Administration System. *Acta Obstet Gynecol Scand* 2007; 86:205–209.
- (15) Palmér S, Blomqvist C, Holmqvist M, ym. Validation of primary and outcome data quality in a Swedish population-based breast cancer quality registry. *BMC Cancer* 2024;24:329.
- (16) Langhoff-Roos J, Krebs L, Klungsoyr K, ym. The Nordic medical birth registers – a potential goldmine for clinical research. *Acta Obstet Gynecol Scand* 2014;93:132–137. doi: 10.1111/aogs.12302
- (17) Pedersen LH, Liew Z. Is man the only animal with ears that cannot move them? Reflections on observational studies in obstetrics and gynecology. *Acta Obstet Gynecol Scand* 2024;103:1478–9. doi: 10.1111/aogs.14925
- (18) Chia KS. Commentary. *Scand J Work Environ Health* 1997;23:152–154.
- (19) Hvidberg MF, Frølich A, Lundstrøm SL. Catalogue of socioeconomic disparities and characteristics of 199+ chronic conditions—A nationwide register-based population study. *PLOS ONE* 2022;30:e0278380. doi: 10.1371/journal.pone.0278380
- (20) Aittomäki A, Martikainen P, Laaksonen M, ym. Household economic resources, labour-market advantage and health problems – A study on causal relationships using prospective register data. *Soc Sci Med* 2012;75:1303–1013. doi: 10.1016/j.socscimed.2012.05.015
- (21) Ng N, Santosa A, Weinehall L, ym. Living alone and mortality among older people in Västerbotten County in Sweden: a survey and

register-based longitudinal study. *BMC Geriatr* 2020;20:7.

doi: 10.1186/s12877-019-1330-9

- (22) Böckerman P, Kortelainen M, Salokangas H, ym. Syövän vaikutukset puolison tuloihin ja työllisyyteen. *Sos Aikakauslehti*. 2023;60:251–262.
- (23) Oyelese Y. Randomized controlled trials: not always the “gold standard” for evidence in obstetrics and gynecology. *Am J Obstet Gynecol* 2024;230:417–425.  
doi: 10.1016/j.ajog.2023.10.015
- (24) Wedenoja S, Soini S, Porkka K, ym. Rekisteritutkimus uhattuna – toisiolakiin liittyvistä ongelmista kohti ratkaisuja. *Suom Lääkärilehti* 2022;77:e32898.
- (25) The European Health Data Space (EHDS). Luettu 5.9.2024. <https://www.european-health-data-space.com/>

SARI RÄISÄNEN

*TtT, dosentti, lehtori*

*Laurea-ammattikorkeakoulu*

KATARIINA LAINE

*MD, PhD, MBA, Senior Researcher*

*Norwegian Research Centre for Women's Health*

*Oslo University Hospital*

*Associate Professor*

*University of Oslo*

Eeva Liikanen

*TtT, dosentti, yliopettaja*

*Tampereen ammattikorkeakoulu*

*Soveltavan tutkimuksen keskus*

Mika Gissler

*FT, VTm, dosentti, tutkimusprofessori*

*Terveyden ja hyvinvoinnin laitos, tieto-osasto*

*Turun yliopisto, Lastenpsykiatrian tutkimuskeskus ja INVEST-lippulaiva*

*Karoliininen instituutti, molekyyli lääketieteen ja -kirurgian laitos*