

Kuinka tukea sosiaali- ja terveystieteiden rekisteritutkimusta?

Rekisteritutkimuksen tukikeskuksen järjestämä lainsäädäntöä koskeva neuvottelupäivä 20.1.2010

Rekisteritutkimuksen tukikeskuksen (ReTki) järjestää joka toinen vuosi neuvottelupäivän, ja tällä kertaa aiheena oli rekisteritutkimusta säätelevät lait. Suomella on ollut ainutlaatuiset mahdollisuudet nousta rekisteritutkimuksen kärkimaaksi. Tutkimustiedon tarve on tiedostettu aina hallitusohjelmaa myöten ja ReTkikin on tunnustettu kansallisesti merkittäväksi tutkimusinfrastruktuuriksi. ReTki syntyi vuonna 2003 Suomen epidemiologian seuran ja Sosiaalilääketieteen yhdistyksen kahta vuotta aiemmin Suomen Akatemialle tekemästä aloitteesta. On siis kulunut noin kymmenen vuotta siitä, kun Hilkka Riihimäki, Ilmo Keskimäki, Markku Kosken- vuo ja molemmat kirjoittajat koontuivat ensimmäisen kerran Työterveyslaitoksen tiloissa valmistelevaan Akatemialle lähtenyttyä aloitetta.

Neuvottelupäivän avannut rekisteritutkimuksen grand old lady Hilkka Riihimäki kuitenkin arveli, että Pohjoismaista ainakin Ruotsi ja Tanska ovat ajaneet Suomen ohitse tutkimuksen tuen, infrastruktuurin ja julkaisujen määrän suhteen. Miten tässä näin kävi ja miten voisimme jatkossa tilannettamme parantaa? Riihimäen mukaan sietää siis provosoivasti kysyä, haluaako Suomi olla rekisteritutkimuksen kärki- maa vai kehitysmä? Halutaanko

olla harkitsevia sijoittajia vai sukaan säästäviä mummoja?

Neuvottelupäivien kutsuvieraina oli Ruotsin kansallisen tietoinfrastruktuurikomitean (DISC Datainfrastrukturkommitté) johtaja Magnus Stenbeck. Komitean tavoitteena on edistää tieteellisessä tutkimuksessa käytettävien aineistojen saatavuutta ja yhteiskäyttöä. Sen budjetti on noin 10 miljoonaa euroa vuodessa, joten hanke ei ole mitään napeilla pelaamista. Vertailun vuoksi mainittakoon, että ReTkin budjetti lasketaan parissasadassa tuhannessa eurossa.

DISC on valmistellut ruotsin- ja englanninkielisenä julkaistun JUDIRaportin (www.disc.vr.se: Rättsliga aspekter- JUDI), jossa kartoitetaan tutkimusrekisterien muodostamisen juridisia esteitä Ruotsissa. Näitä onkin suurten tietokantojen ja tutkimusaineistojen muodostamisen tiellä monia: vaihtelevat ja monesti tiukat tietosuoja- ja arkistointimääräykset sekä eri viranomaisten erilaiset lupakäytännöt. Rekisteritietojen hyödyntämisestä haittaa myös se, ettei olemassa olevista aineistoista eikä niiden tietosisällöstä ole yhteen koottua tietoa. Tähän tarvittaisi keskitetty taho, jolla olisi lakiin perustuva velvoite koota tietoa rekistereistä ja niiden tietosisällöstä. Lisäksi dokumentointi voi olla hyvin- kin puutteellista ja kirjavaa. Kuulostaako tutulta myös Suomessa?

Rekisteriaineiston muodostaminen voi tyssätä jo alkumetreihin. Ruotsin lain mukaan kaikki tutkimukset on arvioitava eettisesti. Toisaalta eettiset toimikunnat eivät voi antaa lupaa tietokantojen perustamiseen, ainoastaan arvioida aineiston tutkimuskäytön eettisyyttä. Lisäksi aineiston muodostamiselle pitää olla tarkasti määriteltävissä

olevat tavoitteet: yleistavoitteita varten rekisteriaineistoja ei voida muodostaa tai edes eettisesti arvioida yleistavoitteisiin väistämättä sisältyvän epätieteellisyyden takia. Lisäksi rekisterinpitäjillä ei välttämättä ole tehokasta infrastruktuuria, joka mahdollistaisi joustavan tietojen luovutuksen ja aineistojen muodostamisen kustannukset voivatkin olla huomattavat. Asioiden korjaamiseksi lainsäädäntöä olisi muutettava. Stenbeck otti esille myös mielenkiintoisen ongelman liittyen tekijänoikeuskysymyksiin, joka juontaa juurensa siitä, ettei tutkimustulosten ja -aineiston ero ole aina mitenkään kristallinkirkas.

Lainsäädännöllisten esteiden lisäksi tekniset ongelmat voivat olla merkittäviä esteitä, etenkin eri rekisterien yhteiskäytössä. DISC:n valmisteilla oleva CODIR-raportti käsittelee erityisesti erilaisia teknisiä ratkaisuja, esimerkiksi epidemiologisessa tutkimuksessa. Integroitujen tietokantojen kantavana ajatuksena on, että online-järjestelmä on tietosuojan kannalta parempi kuin tutkimusaineistojen lähettäminen tutkijoille. Tietokantoja ei yhdistetä fyysisesti, vaan itse aineistot säilyvät rekisteristä vastaavalla laitoksella. Tutkija ei saa koskaan henkilötason aineistoa itselleen, ainoastaan etäyhteyden eri rekisterien yhdistellyistä aineistoista, kuten sikkäläisen Tilastokeskuksen Microdata Online System (MONA) -järjestelmässä. Rekisterien yhteiskäyttöön vaikuttavat lait selvitetään myös laajemmin Pohjoismaiden ministerineuvoston rahoittamassa hankkeessa. Sitä johtaa Magnus Stenbeck ja Suomen edustajana on dosentti Marjut Salokannel Helsingin yliopistosta. Lisävoimia perään-

kuulutettiin. Suomessa ReTki tekee vastaavaa etäkäyttöselvitystä CSC – Tieteen tietotekniikan keskuksen kanssa opetusministeriön rahoituksesta keväeseen 2011 mennessä.

Yleisesti Magnus Stenbeck kannatti, että rekisterinpitäjille maksettaisiin keskitetyistä tutkimusta tukevista varoista korvaukset tutkimusaineistojen muodostamisesta sen sijaan, että rekisterinpitäjät määrittävät omat tutkijoilta perittävät hintansa. Korkeiden aineistonmuodostamiskustannusten ei pitäisi estää tieteellistä tutkimusta. Tähän myös suomalaista rekisteritutkimusta koskevaan maksupolitiikkaan palattiin neuvottelupäivän aikana useaan otteeseen.

ReTkin päällikkö Irma-Leena Notkola esitteli rekisterien toiminnan lakiperustaa ja rekisteritutkimusta sääntelevää lainsäädäntöä. Suomen rekisterijärjestelmä koostuu hallinnollisista rekistereistä sekä erilaisista tilasto-, tutkimus-, suunnittelu- ja valvontarekisteristä. Näitä koskevat eri lainsäädännöt, ja niitä tulkitaan eri tavoin. Rekisteritutkimusta hankaloittavat myös erilaiset luvansaamisprosessit esimerkiksi puhtaassa rekisteritutkimuksessa tai käytettäessä rekistereitä täydentämään muita aineistoja, kuten kyselytutkimustietoja, sairaskertomustietoja tai geenitietoja.

Eri rekisterinpitäjät vastaavat omista luvistaan, joten rekisteritutkija kohtaa varsinaisen laki- ja lupaviidakon. Vaikka ReTki onkin pyrkinyt tasoittamaan tietä (ks. www.rekisteritutkimus.fi), on jäljellä edelleen monia mutkia. Suurimmat ongelmat Notkolan mukaan ovat 1) lakien lukuisuus ja erilaisuus, 2) yleis- ja erityislakien väliset erot ja 3) rekisteritutkimusta vaikeuttava tilastolaki.

Rekisteritutkimusta koskevia yleislakeja ovat henkilötietolaki ja julkisuuslaki, joita on aina noudatettava. Suurin osa rekisterilainsäädännöistä on erityislakeja, kuten

laki terveydenhuollon valtakunnallisista henkilörekistereistä ja tilastolaki. Yleislaeista voidaan erityislaeissa joko poiketa (sijasta) tai niitä voidaan täydentää (lisäksi). Eri lait ovat välillä keskenään ristiriitaisia, mikä aiheuttaa hämmennystä. Esimerkiksi laki terveydenhuollon valtakunnallisista henkilörekistereistä on kolmatta vuosikymmentä vanha, ja se kaippaa kipeästi päivitystä.

Tilastolain nykyisen tulkinnan mukaan tutkimukseen käytettävistä rekisteriaineistosta ei saa tunnistaa yksilöä edes epäsuorasti. Näin olleen tarvitaan täydellistä anonymisointia, eikä kokonaisaineistoja voida luovuttaa. Tilastokeskuksen aineistoista voidaan henkilötunnuksellisia luovuttaa ainoastaan kuolemansyytietoja sekä ikä-, sukupuoli-, ammatti- ja koulutustietoja. Tutkijan kannalta aineiston laatu on heikompi ja aineiston saaminen viipyy anonymisoinnin ja karkeistamisen vuoksi. Laskun maksaa tutkija korkeina aineistokustannuksina. Tilastolaki voi estää tutkimuksen kokonaan. Esimerkiksi Tilastokeskuksen ja THL:n sosiaalirekistereiden aineistoja ei voida yhdistää tutkimuksissa, sillä nämä tilastoviranomaiset voivat luovuttaa henkilötiedot sisältäviä aineistoja toisilleen ainoastaan tilastointia, ei tutkimusta varten.

EU:n tilastotoimintaa koskeva asetus on hyväksytty vuonna 2009. Sen mukaan tutkijalle voidaan luovuttaa aineisto, josta ainoastaan suorat tunnisteen poistetaan. Keskustelussa Tilastokeskuksen lakimies Johanna Rantanen kertoi, että EU:n tilastoasetus on mahdollistava, ei velvoittava. Periaatteessa se koskee vain niitä tilastoja, jotka määrittellään Eurostatin tilasto-ohjelmassa ja sen vuosiohjelmissa. Epäselväksi jäi, kattaako tämä vain EU:lle toimitettavat tiedot, vaan myös kaikki tilastojen saamiseksi tarvittavat aineistot. Roope Uusi-

talo (Valtion taloudellinen tutkimuskeskus) nosti esille esimerkkinä kansantalouden tilinpitotiedot, joita käytetään EU:lle toimitettavien tietojen muodostamiseen, mutta jotka samalla sisältävät paljon toimitettavia tietoja yksityiskohtaisempaa aineistoa. Yleisö oli hyvin yksimielinen siitä, että kansallista tilastolakiä olisi muutettava EU-tilastolain linjan mukaiseksi. Päivän kohokohdista oli, kun Johanna Rantanen kertoi Tilastokeskuksen jo esittäneen valtionvarainministeriölle tilastolain uudistamista.

Tietosuojalainsäädännön lisäksi valtion maksuperustelaki hankaloittaa rekisteritutkimusta. Epäloogisesti Tilastokeskuksen on noudatettava liiketaloudellisia periaatteita aineistojen muodostamisessa, mutta Väestörekisterikeskuksen julkisoikeudellisia periaatteita. Etenkin kustannusten määrittelemisen edelleen tapausmäärän perusteella herättää ihmetystä, koska tietokoneajo ja siihen kuluva aika on lähes riippumaton tapausmäärästä. Ehkä tätäkin lakia olisi päivitettävä tälle vuosituuhannelle? Kommentissa Väestörekisterikeskuksen edustaja Riitta Haggren kuitenkin huomautti, että tiedon irrottaminen voi olla vaikeaa ja kallista, koska palvelut ostetaan ulkoa. Itse asiassa tutkijoille tietoja annetaan halvemmallalla, joten niiltä osin toiminta on tyypillisesti alikatteista.

Irma-Leena Notkola visioi esityksensä päätteeksi uutta tutkimuslakia, joka helpottaisi rekisteritutkimusten luparumbaa. Kun tähän vielä lisättäisiin luvan hankkiminen yhden luukun periaatteella ja sähköiset lupaprosessit sekä mahdollisuus aineistojen etäkäyttöön, suurin osa nykyisistä rekisteritutkimuksen esteistä olisi jo karsittu. Hän vilautti myös visiota etäkäyttäjärjestelmälle, joka kulkee työnimellä MIDRAS (Micro Data Remote Access System) ja josta on käynnissä jo opetusministeriön rahoittama val-

misteluhanke. Aiemmin on myös esitetty idea ”koodipankista”. Etäkäyttäjärjestelmän käyttöluvan saamisen edellytyksenä tutkijoilla olisi velvollisuus julkistaa aineiston muokkaamiseen ja analysointiin käyttämänsä lähdekoodit ja skriptit. Näin muiden tutkijoiden ei tarvitsisi keksiä pyörää aina uudestaan, kun tyyppillisiin aineiston (esi) käsittelytoimenpiteisiin löytyisi jo sopivia standardipohjia. Etuna olisi myös tutkimusten toistettavuuden ja oikeellisuuden konkreettinen todentamismahdollisuus, joka tältä osin parantaisi tutkimuksen objektiivista laatua.

Neuvottelupäivän tutkijapuheenvuoron piti tutkimusjohtaja Roope Uusitalo. Aivan aluksi hän muistutti, että tutkimuslaitoksissa suhtaudutaan tietosuojakysymyksiin vakavasti. Tutkijoilla ei ole mitään tarvetta pyrkiä tunnistamaan havaintoyksiköjä tai yksittäisiä henkilöitä tai raportoida niitä tutkimusraporteissa. Tietoaaineistojen käytön lisääminen vaatii niiden saatavuuden parantamista ja hinnoittelun kohtuullistamista. Henkilötietolaki on tutkimusmyönteinen, mutta tilastolaki ei. Sen selventäminen ja muuttaminen tutkijamyönteisemmäksi olisi ensisijaisen tärkeää. Aineistokustannukset eivät rekisteritutkimusta kaada, vaan kohtuulliset kustannukset voidaan tutkijalta laskea. Olemassa olevien tutkimusaineistojen saatavuutta hän toivoi parannettavaksi esimerkiksi lisäämällä niitä kansallisiin tutkimusarkistoihin, kuten Tampereella toimivaan Yhteiskunnalliseen tietoarkistoon (FSD).

Roope Uusitalo päätti esityksensä esimerkeillä onnistuneista ja vähemmän onnistuneista rekisteritutkimuksista. Työttömyysturvaa koskevan tutkimuksen aineiston muodostaminen Työministeriön, Vakuutusvalvontaviraston ja Eläketurvakeskuksen rekisteritiedoista hoitui todella nopeasti ja edullisesti.

Tämä tuskin olisi onnistunut, jos tutkijat eivät olisi tunteneet näitä organisaatioita ja niiden toimintatapoja entuudestaan. Toisaalta maahanmuuttajien kotouttamislainsäädäntöä arvioivassa tutkimuksessa Tilastokeskus teki kalliin aineiston, mutta luovutti lopulta vain 15 prosenttia otoksen tutkijoille. Kaikkien maahanmuuttajien osalta kotouttamislainsäädäntö näytti parantaneen maahanmuuttajien taloudellista asemaa ja työllisyyttä. Liian pieni otos esti eri maahanmuuttajaryhmien tarkemman analysoinnin, joten paremmalla lainsäädännöllä tutkimustulokset olisivat voineet olla paljon yksityiskohtaisempia ja tarkempia. Paras esimerkki oli kuitenkin se, että erään työvoimapolitiikan kokeilun yhteyteen säädettiin laki, joka velvoitti kokeilun vaikutusten tutkimiseen ja verohallinnon oli luovutettava kokeilun arviointia varten tarvittavat tiedot ilman kustannuksia. Tämäntyyppistä käytäntöä voisi käyttää useimmin.

Tilaisuuteen oli pyydetty kolme kommenttipuheenvuoroa. Tietosuojavaltuutettu Reijo Aarnio muistutti, ettei tietosuojalainsäädäntö ole kompastuskivi. Kun osa rekisterinpitäjistä on nettobudjetoituja, joutuvat ne laskuttamaan rekisteriaineistojen tekemisen kustannukset. Näin yllättäen valtiovarainministeriö johtaa rekisteritutkimuspolitiikkaa.

Reijo Aarnio myönsi, että nykyisen lainsäädännön puutteet ovat hyvin tiedossa, mutta uuden lainsäädännön tekemistä haittaa juristien aikapula. Kiireen vuoksi yhteiskunnallisten sidosryhmien lainvalmistelijoille toimittamat kommentit saattavat jäädä huomioimatta. Tutkimusta hankaloittavat myös monimutkaiset hallinnon rakenteet. Jos Google on paras tiedonetsintäkeino, ei viranomaisen taatusti noudata julkisuuslain henkeä.

Reijo Aarniolla oli monia ehdo-

tuksia. Ensiksi pitäisi selvittää peruskäsitteet. Edelleenkin juristeille on epäselvää, mitä tieteellisellä tutkimuksella tarkoitetaan. Kipupisteanalyysilla pitäisi tunnistaa ne keinot, kuinka tutkijoita voidaan palvella paremmin. Esimerkiksi anonymisointia ei osata eikä ulkopuolisia anonymisointipalveluita ole olemassa. Tällaisten palveluiden pitäisi myös kattaa eri aineistojen linkkaukset, ainakin siihen asti, kunnes integroidut tietokannat saadaan toimimaan. Lisäksi marssijärjestys saattaa olla epäselvä: kuka voi ja haluaa laittaa lainmuutokset vireille? Aarnion mukaan asia on vastuutettava sopivalle taholle. Pitää tehdä jotain, eikä vain puhua!

Tilastokeskuksen lakimies Johanna Rantanen kertoi tilastolain taustasta. Aiemmin Tilastokeskus joutui kysymään luvat kaikkien perusaineistojen rekisterinpitäjältä erikseen. Vuoden 1994 lain mukaan lupia ei enää tarvittu kysyä, eikä vuonna 2004 uusitun lain valmistelussa lainsäädännöstä tutkimukselle aiheutuneita ongelmia nostettu esille esiin. Keskustelussa Roope Uusitalo kuitenkin totesi ainakin VATT:n antaneen lausunnon koskien välilistä tunnistusta, mutta ilmeisesti sen painoarvo oli ollut liian pieni.

Tilastokeskuksessa on ollut oma etäkäyttöhankkeen pilotti, joka päättyi lokakuussa 2009. Sitä jatketaan käytettävissä olevin, arvattavasti liian vähäisin resurssein. Kolme tutkimuslaitosta käyttää jo etäyhteysmahdollisuutta. Itse aineisto on siis edelleen ainoastaan Tilastokeskuksessa ja tutkijat saavat vain lopputulokset käyttöönsä. Roope Uusitalo mainitsi päässeensä tutustumaan pilottiin, joka oli vaikuttanut hyvältä, kunhan sinne saataisiin lisää aineistoja.

Marjut Salokannel Helsingin yliopistosta kertoi vuoden 2010 loppuun asti toimivasta tutkimuksen tietoaaineistojen saatavuutta käsittelevästä poikkihallinnollisesta

ohjausryhmästä, jonka puheenjohtajana toimii Ilkka Niiniluoto. Ryhmä nostanee esille asioita, joista jo muutkin neuvottelupäivän esiintyjät olivat huolissaan: 1) Henkilötietolain ja tilastolain välinen suhde on ongelmallinen. Muualla kuin Suomessa henkilötietolakia on tulkittu tilastolakia tiukemmin. Käytännöt on selvitettävä ja yhtenäistettävä tutkimukseen. 2) Maksuperustelaki on vanhentunut, mikä vaikuttaa aineistojen saatavuuteen hinnoittelun kautta. Yleisperiaatteen mukaan tutkijoiden pitäisi saada aineistot alhaisin kustannuksin, jotka kattavat vain irrottamisesta aiheutuneet marginaalikustannukset. 3) Tietoa-aineistojen lisensointi ja tekijänoikeuslainsäädäntö kaipaavat päivittämistä. Tekijänoikeudesta vapaita teoksia ei saisi lisensoida, vaan niiden täytyisi olla vapaasti saatavissa. Tekijänoikeuslainsäädännön mukaan suojeltujen tietojen käyttäminen tieteelliseen tutkimukseen pitäisi olla mahdollista, kunhan lähde ja tekijän nimi mainitaan.

Yleiskeskustelussa Magnus Stenbeck huomautti, että Ruotsin ja Suomen ongelmat näyttävät olevan samanlaisia. Yksi ydinkysymys on, voidaananko pysyviä rekistereitä perustaa ainoastaan tutkimustarkoituksiin. Reijo Aarnion mukaan on selvitettävä, voiko pysyvän tiedonkeruun perustana olla tieteellinen tutkimus. Valistunut arvaus oli, ettei tällainen mahdollisuus ole mahdollista nykyllä lainsäädännön puitteissa.

Marja Pirttivaaran (SITRA) mukaan on tärkeää, että sähköistä

tiedonkeruuta koskeva lainsäädäntö on yhteneväinen muiden tiedonkeruumuotojen kanssa. Pirttivaara kaipasi myös anonymisointipalvelua. Mia Bengtström Lääketeollisuus ry:stä kertoi, kuinka sähköisiä järjestelmiä koskevassa keskustelussa voidaan anonymisoinnissa käyttää niin sanottua kaksoiskoodausta. Eri tietolähteiden tunnistet salataan eri avaimilla, mutta ne voidaan tarvittaessa yhdistää. Pirttivaaran mukaan myös rekisteriselosteiden sisältöä pitäisi laajentaa siten, että niihin voisi viitata tieteellisissä tutkimuksissa. Kenellekään ei jäisi epäselväksi, mistä rekisteristä ja tiedoista oikein on kyse ja millaista aineiston laatu on. Ehkä tarvitsemme Tilastokeskuksen Suomen virallisen tilaston (SVT) laatu-kriteerien rinnalle Suomen virallisen rekisterin (SVR) käsitteen?

Riitta-Sisko Koskela (TTL) muistutti, että edelleen tarvitaan tutkijoiden ohjausta ja opetusta, koska erilaisia sudenkuoppia on enemmän kuin tarpeeksi. Koskela korosti myös tutkimuksen suunnittelun tärkeyttä, koska hyvin suunnitellussa tutkimuksessa pienempikin otos voi aivan hyvin riittää vastauksen saamiseksi tutkimuskysymyksen. Myös Piia-Liisa Heiliö (STM) toivoi, että kokeneet tutkijat jakaisivat auliisti tietämystään jo suunnitteluvaiheessa. Myös rekisteriselosteiden pitäisi olla parempia. Nykyisellään turhan paljon lupaviranomaisten ajasta kuluu saatavilla olevista aineistoista kertomiseen. Tutkimuslupienkin käsittelyaika on pitkä, koska hakemukset ovat usein

puutteellisia. Helena Kääriäinen (THL) ehdotti, että rekisteriviranomaiset julkistaisivat keskimääräiset lupakäsittelyajat. Tässä pitäisi kuitenkin erotella viranomaisesta ja hakijasta johtuvat viivästyks.

Hilkka Riihimäki kokosi päivän aikana todetut asiat ja niiden jatkotoimenpiteet. Ongelmat on tunnistettu jo kauan aikaa sitten, mutta koordinaatio puuttuu eikä toimintaa ole saatu aikaiseksi. Yksittäisellä tutkijalla ei ole juurikaan mahdollisuutta vaikuttaa asioihin. Rekisteritutkimuksen tukikeskus voi omalla toiminnallaan edistää asioita, mutta ehdotusten pitäisi olla konkreettisia, eikä ”innovaation” tai ”infrastruktuurin” kaltaisia abstraktioita. Tuomas Parkkari (Tutkimus- ja innovaationeuvosto) kertoi, että vuonna 2010 valmistuva tutkimus- ja innovaationeuvoston linjaraportti pyrkii korjaamaan tätä epäkohtaa. Tutkimuksen tärkeä merkitys nostetaan raportista selvästi esille. Hän toivoi, että neuvostolle toimitettaisiin rakentavia ehdotuksia mm. rekisteritutkimuksen edistämisestä. Myös Niiniluodon työryhmän selvitys valmistuu vuoden loppuun mennessä. Kyseessä ovat isot asiat ja keskustelu on avattu!

Neuvottelupäivän esitykset ovat saatavilla Rekisteritutkimuksen tukikeskuksen verkkosivuilta (www.rekisteritutkimus.fi: Toiminta – Muu toiminta).

MIKA GISSLER
REIJO SUND