

Kriminologisen rekisteritutkimuksen kehityssuuntia Pohjoismaissa

Veli Verkko –luento 14.12.2020

MIKKO AALTONEN

Rikollisuutta koskevien rekisteriaineistojen tutkimuskäyttö on lisääntynyt 2000-luvulla voimakkaasti Pohjoismaissa ja myös Suomessa. Kriminologisen tutkimuksen mahdollisuudet ovat parantuneet rekisterilinkkausaineistojen aikaisempaa paremman saatavuuden myötä, ja erityisesti Tilastokeskuksen etäkäyttöjärjestelmä FIONA:n perustaminen on ollut tärkeä uudistus. Rekisteriaineistoja on hyödynnetty viime vuosina sekä rikollisuuden syiden että rikosseuraamusten vaikutusten pohjoismaisessa tutkimuksessa varsin monipuolisesti. Yksilötasoisilla pitkittäisaineistoilla päästään kiinni huomattavasti aggregaattitilastoja tarkempaan rikollisuuden muutosten analyysiin ja niitä voidaan hyödyntää aikaisempaa uskottavammassa kausaalivaikutusten tutkimuksessa. Luonnollisten koeasetelmien ja kvasikokeellisten tutkimusmenetelmien kehitys on hyödyttänyt rekisteritutkimusta selvästi ja nostanut osaltaan rekisteriaineistojen käyttöarvoa.

Kirjoitus perustuu 14.12.2020 pidettyyn Veli Verkko -luentoon.

Rikollisuutta koskevilla virallistilastoilla on tärkeä rooli kriminologisen tutkimuksen historiassa. Poliisin tietoon tulleita rikoksia, tuomioita ja vankilukua koskevat tilastot muodostavat tänäkin päivänä keskeisen osan määrällistä tutkimusta tekevien kriminologien työkalupakkia. Kriminologin keskeiseen ammattitaitoon kuuluu näiden tilastojen luotettavuutta – sekä validiteettia että reliabiliteettia – koskevien kysymysten kriittinen pohdinta.

Edellä mainitut virallistilastot laaditaan tänä päivänä viranomaisten operatiivisten tietojärjestelmien tietoja käyttäen. Tilastollisessa tutkimuksessa hyödynnetään yleensä tietojärjestelmissä valmiiksi numeerisessa tai muuten ennalta määrättyssä muodossa olevia tietoja eli niin sanottua rakenteista tietoa. Li-

säksi hallinnolliset rekisteriaineistot (”administrative data”, Connelly ym., 2016) sisältävät asiakirja-aineistoa, joista esimerkiksi esitutkintapöytäkirjat, haastehakemukset ja tuomiot ovat tuttuja laadullista tutkimusta tekeville. Kestohaasteena on se, että asiakirjat ja tietojärjestelmien vapaatekstikentät sisältävät paljon sellaista tietoa, joka olisi kiinnostavaa myös määrällisessä tarkastelussa, mutta tiedon muuntaminen rakenteiseen muotoon ”koodaamalla” vaatii paljon työtä.

Tässä puheenvuorossa tarkoitan kriminologisella rekisteritutkimuksella sellaista tutkimusta, jossa rikollisuutta, rangaistuksia tai täytäntöönpanoa kuvaaviin tietoihin yhdistetään yksilötasolla tietoja muista rekisteriaineistoista. ”Puhtaan” rekisteritutkimuksen lisäksi Suomessa ja maailmalla on myös tehty paljon pitkittäistutkimusta aineistoilla, joissa

mm. kyselyin ja kliinisin tutkimuksin kerättyyn tietoon on liitetty rekisteripohjaista tietoa esimerkiksi poliisin tietoon tulleista rikoksista (Elonheimo ym., 2014).

Olen itse tehnyt erilaisiin rekisteriaineistojen yhdistelmiin perustuvaa tutkimusta vuodesta 2008 lähtien. Tämän verrattain lyhyen ajanjakson kuluessa kriminologisen rekisteritutkimuksen määrä on lisääntynyt voimakkaasti, ja tuolloin vielä sumuiselta vaikuttanut lähestymistapa rikollisuuden tutkimiseen tuntuu nyt itsestään selvältä. Rekisterilinkkausaineistot ovat nousseet erityisesti pohjoismaissa keskeiseksi kriminologisen tutkimuksen tietolähteeksi. Tämä maantieteellinen painotus ei ole yllättävää sikäli, että pohjoismaissa tilastotuotanto on jo pitkään nojannut vahvasti rekisteriaineistoihin (Tilastokeskus, 2004). Pohjoismaiden lisäksi Hollanti on ollut edelläkävijä rekisteripohjaisessa tutkimuksessa (Blokland & Nieuwbeerta, 2005).

Kriminologisen rekisteritutkimuksen lisääntyä on perusteltua kysyä, millaista uutta tietoa rikollisuudesta ja sen kontrollista on rekisteriaineistoilla saatu? Esitän seuraavassa (väistämättä subjektiivisen) näkemyksen siitä, minkälaisissa tutkimusasetelmissä rekisteritutkimus on vahvimmillaan, ja millaisiin tutkimuskysymyksiin rekisteriaineistoilla pystyy vastaamaan.¹ Kyseessä ei ole systemaattinen katsaus kaikista pohjoismaisista tai edes suomalaisista tutkimuksista, vaan pikemminkin pyrkimys kuvata eräitä keskeisiä lähestymistapoja ja kekseliäitä tutkimusasetelmia esimerkkien valossa. Pääosassa ovat tutkimukset, jotka ovat nähdäk-

seni onnistuneet uskottavasti tuottamaan tietoa vaikutuksista pelkän yhteisvaihtelun sijasta. Rekisteriaineistot eivät vastaa läheskään kaikkiin kriminologisen tutkimuksen tietotarpeisiin, ja voivat pahimmillaan tuottaa väärintulkittuina suorastaan harhaanjohtavaa tietoa. Vahvuusalueellaan rekisteriaineistot ovat kuitenkin ylittämättömiä, ja mahdollistavat varsin monipuolisen empiirisen tutkimuksen.

Suomalainen kriminologinen rekisteritutkimus 2000-luvulla

Olenaisimpana edellytyksenä laadukkaan rekisteritutkimuksen tekemiselle on luotettava väestökirjanpito ja henkilöiden yksilöitävyys läpi rekisterien samalla tunnisteella. Yksilöinnin mahdollistava henkilötunnus otettiin Suomessa käyttöön 1960-luvulla, väestökirjanpidon perustana toimiva Digi- ja väestötietoviraston ylläpitämä Väestötietojärjestelmä puolestaan vuonna 1969 (Tilastokeskus, 2004). Laadukkaan väestökirjanpidon avulla voidaan lisäksi määrittää henkilöiden perhesuhteet, ja luoda aineistoja, joissa ovat mukana sekä vanhemmat että lapset, mahdollisesti jopa useampia sukupolvia. Rekisteriaineistojen keskeinen vahvuus onkin siten laajojen yksilötasojen pitkittäisaineistojen *retrospektiivisessä* rakentamisessa. Erityisesti terveysrekistereillä on Suomessa pitkä historia tutkimuskäytössä (Sihvonen & Gissler, 2005).

2000-luvun alussa tehdyt rekisteriaineistoja hyödyntäneet tutkimukset laadittiin pääsääntöisesti yhteishankkeina Tilastokeskuksen

¹ Hyviä johdatuksia rekisteritutkimuksen perusperiaatteisiin, mahdollisuuksiin ja sudenkuoppiin ovat mm. Lyngstad & Skardhamar, 2011 ja Connelly ym., 2016.

kanssa. Tuohon aikaan Tilastokeskuksen hallussa olevia aineistoja analysoitiin Tilastokeskuksen tiloissa ns. tutkijalaboratoriossa, eikä aineistoja voinut luovuttaa talon ulkopuolelle. Muun muassa Päivi Honkatukia (2001), Aarne Kinnunen (2001) ja Heini Kainulainen (2004) hyödynsivät rikosilmoitus- ja tuomioaineistoja raiskausrikoksia ja huumausainerikoksia koskeneissa tutkimuksissaan.

Vähitellen rikostietoja sisältäviä rekisteriaineistoja alettiin muodostaa myös Tilastokeskuksen ulkopuolella. Sektoritutkimuslaitoksissa, joilla oli itsellään hallussa valmiiksi eräitä rekisteriaineistoja, yhdistettiin muun muassa Oikeuspoliittisen tutkimuslaitoksen Suomalaisen rikoskäyttäytymisen riskitekijät-aineisto (Aaltonen, 2013) sekä Terveyden ja hyvinvoinnin laitoksen 1987-syntymäkohorttiaineisto (Paananen ym., 2012). Näistä erityisesti 1987-syntymäkohorttiaineisto (sekä sitä seurannut 1997-kohorttiaineisto) perustuu hyvin monipuoliseen aineistolinkkaukseen, ja mahdollistaa edelleen monipuolisen tutkimuksen. Näihin aineistoihin yhdistetyt rikollisuutta koskevat rekisteritiedot haettiin Oikeusrekisterikeskukselta tai poliisilta. Pääosa Tilastokeskuksen rekisteriaineistoista jäi kuitenkin näiden sektoritutkimuslaitoksissa yhdistettyjen aineistojen ulkopuolelle, sillä niitä ei tietosuojasyistä luovutettu talon ulkopuolelle (ks. esim. Sihvonen & Gissler, 2005). Joitakin poikkeuksia tähän kuitenkin tehtiin (Savolainen & Sirén, 2013).

Eriyisesti Tilastokeskuksen rekisteriaineistojen käyttö on helpottunut merkittävästi tilastolain muutoksen (HE 154/2012 vp) ja etäkäyttöjärjestelmä FIONA:n käyttöönoton

myötä viime vuosina. Lakimuutoksen keskeisenä tavoitteena oli edistää rekisteritietojen nykyistä laajempaa käyttöä tieteellisessä tutkimuksessa. Suomi seurasi muiden pohjoismaiden esimerkkiä perustamalla rekisteritutkimuksen käyttöön etäkäyttöympäristön, jossa aiemmin ainoastaan tutkijalaboratoriossa käytössä olleita Tilastokeskuksen aineistoja on voitu luovuttaa pseudonymisoina etäkäyttöön. Etäkäyttöjärjestelmää ylläpitää Tieteen tietotekniikan keskus CSC, jonka palvelimilla yhdistetyt aineistot myös sijaitsevat.

Tilastokeskuksen Rikos- ja pakkokeinotilaston ja Syytetyt, tuomitut ja rangaistukset -tilaston pohjana olevien rekisteritietojen ajallinen kattavuus riittää varsin moniin tutkimustarpeisiin. Poliisin tietoon tulleista rikoksista on saatavilla yksilötasoista tietoa vuodesta 1996 eteenpäin, ja käräjäoikeuksien tuomioista jopa vuodesta 1977 eteenpäin. Poliisin tietoon tulleita rikoksia tai tuomittuja rangaistuksia koskevia aineistoja ei tämän artikkelin kirjoitushetkellä ollut saatavina valmisaineistona, vaan ne räätälöidään Tilastokeskuksen toimesta kunkin tutkimusprojektin tarpeisiin. Valmisaineistoina saatavilla olevista rekisteritiedoista saa ajantasaisen tiedon Tilastokeskuksen verkkosivuilta².

Tutkimuksen painopisteitä Pohjoismaissa

Rekisteritutkimuksen lupakäytäntö näyttää kulkeneen kaikissa Pohjoismaissa laajemmat aineistot sallivaan suuntaan, ja monet tutkimukset hyödyntävätkin nykyään otosten sijaan kokonaisaineistoja, esim. kokonaisia syntymäkohortteja. Tilastollisen voiman paraneamisen lisäksi kokonaisaineistot ovat hyö-

² Taika-tutkimusaineistokatalogi, <https://taika.stat.fi/>

dyllisiä sikäli, että aiemmin pelkkiin aggregoituihin vuositilastoihin perustuneita analyyseja voidaan nyt tehdä yksilötason aineistoilla säilyttäen vertailukelpoisuus julkaistuihin tilastoihin. Tällaisilla aineistoilla on paljon mahdollisuuksia yksinkertaisen mutta potentiaalisesti erittäin informatiivisen kuvailevan analyysin saralla.

Yksi keskeinen kokonaisaineistojen käyttötapana on rikollisuuden ajallisen muutoksen tarkastelu alaryhmissä. Sen sijaan että tyytyisimme tarkastelemaan pelkästään rikollisuuden kokonaismäärän muutoksia, populaatiotason pitkittäisaineistoilla voidaan esimerkiksi tarkastella sitä, missä väestöryhmissä rikollisuuden lasku on ollut voimakkainta (Sivertsson ym., 2019), tai sitä, selittykö nuorisirikollisuuden määrän lasku ensisijaisesti rikoksia tekevien henkilöiden määrän laskulla vai rikoksia tekevien nuorten tekemien rikosten määrän laskulla (Andersen ym., 2016). Vaikka tällainen analyysitapa ei suoraan kerro vastausta siihen, mitkä tekijät selittävät rikollisuuden määrän muutoksia, antavat ne eväitä jatkotutkimukselle sen suhteen, mistä mahdollisia syitä kannattaisi lähteä etsimään, ja mihin väestöryhmään voimakkain muutos paikallistuu.

Yksilötasoisten paneelianeistojen avulla on puolestaan mahdollista tarkastella rikoskäyttäytymisen määrää ja muutosta läpi elämänsä. Ns. rikosuritutkimusta on tehty sekä kyselypohjaisilla pitkittäisaineistoilla että tuomioita tai rikosepäilyjä mittaavilla rekisteriaineistoilla. Pioneeritutkimuksena Pohjoismaista voidaan mainita Britta Kyvsgaardin *The Criminal Career* (2002). Ryhmäperustaiset kasvukäyrämallit (”group-based trajectory modelling”, Nagin, 2005) olivat rikosuramallinnuksessa suosittuja 2000-luvun alkupuolella, mutta vaikuttaa siltä, että tämän lähestymistavan suosio olisi laantunut viime

vuosina. Yhtenä syynä tähän saattaa olla menetelmää kohtaan esitetty kritiikki (Skardhamar, 2010). Rikosuritutkimus ei ole kuitenkaan kadonnut mihinkään. Pitkittäisaineistot mahdollistavat pääsääntöisesti poikkeusleikkausaineistoja uskottavamman rikollisuuden syiden analyysin, ja kuvaankin seuraavassa eräitä keskeisiä lähestymistapoja, joilla kausaalivaikutuksiin pyritään pääsemään kiinni.

Kenties suoraviivaisin tapa hyödyntää pitkittäisaineistoa syytutkimuksessa on siirtyä *yksilöidenvälisestä* analyysistä *yksilönsisäiseen* analyysin käyttämällä kiinteiden vaikutusten (”fixed effects”) regressiomalleja (Allison, 2009). Olemme itse hyödyntäneet näitä malleja työmarkkina-aseman (Ramakers ym., 2020) ja velkaongelmien (Aaltonen ym., 2016) rikosvaikutuksia koskevassa analyysissä. Perusajatuksena näissä asetelmissä on, että esimerkiksi työttömien ja työssäkäyvien henkilöiden rikostasojen vertailun sijaan vertaammekin saman henkilön rikostasoa näissä kahdessa (tai useammassa) vaihtoehtoisessa tilassa eri ajankohtina. Näin voimme epäsuorasti ottaa huomioon staattiset riskitekijät, jotka mahdollisesti selittävät sekä rikoskäyttäytymistä että valikoitumista eri työmarkkina-asemiin. Intuitiivisesti mallin voi ymmärtää siten, että rikoskäyttäytymiseen vaikuttava riskitekijä, joka pysyy ajan yli vakiona, ei voi yksilötasolla selittää eroa työttömänä ja työllisenä. Näiden mallien heikkoutena ovat mahdolliset vastavuoroiset yhteydet (rikollisuus aiheuttaakin työttömyyttä) sekä havaitsemattomat ajassa muuttuvat tekijät. Siksi näiden mallien estimaateille ei lähtökohtaisesti voi antaa kausaalitulkintaa. Tästä huolimatta kiinteiden vaikutusten mallia voi pitää hyvänä ”happotestinä” ajassa muuttuvien tekijöiden yhteyksien tarkastelussa. Ei ole harvinaista, että vahvakin yksilöidenvälisessä analyysissä havaittu muuttu-

jien välinen yhteys kutistuu lähes olemattomiin yksilönsisäisessä mallissa (Aaltonen ym., 2016).

Kiinteiden vaikutusten regressiomalleja voidaan hyödyntää myös erilaisissa geneettisesti informoiduissa, perhe- ja sukulaisuussuhteiden hyödyntämiseen perustuvissa tutkimusasetelmissä (D’Onofrio ym., 2020). Perheasetelmia hyödyntävää rekisteritutkimusta on tehty erityisesti Ruotsissa Karoliinisessa instituutissa varsin monipuolisesti (esim. Frisell, 2012). Vaikka perheasetelmat usein yhdistetään käyttäytymisgeneettiseen tutkimukseen, tarjoavat ne paljon mahdollisuuksia myös ympäristövaikutusten tutkimiselle. Koska perheasetelmat vaativat tyypillisesti suuren aineistokoon saavuttaakseen riittävän tilastollisen voiman erityisesti harvinaisten altisteiden ja vasteiden tutkimuksessa (D’Onofrio ym., 2020, s. 40), ilman rekisterilinkkausaineistoja tällaisen tutkimuksen tekeminen olisi varsin vaikeaa. Rikollisuutta koskevia, perheasetelmaa hyödyntäviä analyyseja on tehty muun muassa lapsuuden perheen tulotason (Sariaslan ym., 2014) ja vankilasta vapautumisen jälkeisen asuinalueen ominaisuuksien vaikutuksista (Sariaslan ym., 2017; ks. myös Latvala, 2020).

Yksilönsisäisiä analyyseja ja perheasetelmia voi ajatella yleisempinä lähestymistapoina, joilla voi tutkia varsin monenlaisia kysymyksiä paneeli- ja perheaineistoilla. Näiden lisäksi kriminologiassa on havaittavissa kasvava kiinnostusta ns. luonnollisten kokeiden hyödyntämiseen kvasikokeellisina tutkimusasetelmina. Varsin usein luonnolliset koeasetelmat nousevat erilaisista politiikka- ja lainsäädäntömuutoksesta tai hallinnollista säännöistä ja käytännöistä, jotka synnyttävät satunnaista vaihtelua johonkin kiinnostavaan rikollisuuden riskitekijään tai vaikkapa rikosseuraamukseen. Tutkimusinstituutioiden ta-

solla tarkastellen erityisesti Tanskan Rockwool Foundationin tutkimusyksiköllä on ollut vahva panos politiikka- ja lainsäädäntömuutoksista nousevien luonnollisten koeasetelmien hyödyntämisessä kriminologian kentällä (esim. Andersen ym., 2019; Anker ym., 2020).

Yksi keskeinen painopistealue luonnollisia koeasetelmia käyttävässä tutkimuksessa on ollut rikosseuraamusten vaikutusten analyysi. Useammassa pohjoismaisessa tutkimuksessa (Andersen & Andersen, 2014; Andersen, 2015; Andersen & Telle, 2019; Fallesen & Andersen, 2017; Larsen, 2017) on tarkastelu avoseuraamusten yhdyskuntapalvelun tai valvontarangaistuksen – vaikutuksia uusintarikollisuuteen ja sosiaaliseen integraatioon (työttömyys, riippuvuus sosiaaliturvasta, koulutus, perhe-elämä) ehdottomaan vankeusrangaistukseen verrattuna. Näissä tutkimuksissa on tyypillisesti hyödynnetty uuden seuraamusmuodon käyttöönottilanetta, ja vertailtu uuteen seuraamusmuotoon tuomittuja ”koeryhmäläisiä” ehdotonta vankeutta suorittaneeseen verrokkiryhmään. Näiden lisäksi on tutkittu muun muassa vankeusrangaistuksen pituuden vaikutuksia (Landersø, 2015). Muista rikollisuuden kontrolliin liittyvistä tutkimuksista on syytä mainita erittäin kiinnostava tanskalaistutkimus DNA-näyteteksterin käytön laajentamisesta (Anker ym., 2020).

Erilaisten yhteiskunnallisten muutosten, shokkien ja politiikkareformien rikosvaikutuksen tutkimiseen kvasikokeellisessa viitekehyksessä on paljon mahdollisuuksia. Tanskassa (Bennett & Ouazad, 2019) ja Norjassa (Rege ym., 2019) on tehty tutkimusta mas sairittaisien yhteydessä työttömäksi joutuneiden rikollisuudesta, Tanskassa on puolestaan saatu tutkimusnäyttöä työvoimapolitiittisten aktivointitoimien rikollisuutta vähentävästä vaikutuksesta (Fallesen ym.,

2018; Andersen, 2012). Ehkä dramaattisin viimeaikainen sosiaalipoliittisten toimien rikosvaikutuksia koskeva tutkimus käsittelee Tanskassa vuonna 2002 tehtyä reformia ("Starthjælp", Andersen ym., 2019), jonka seurauksena turvapaikan saaneiden henkilöiden sosiaaliturvaa leikattiin merkittävästi. Vaikka reformi lisäsi alkuun työllisyyttä, sen havaittiin samalla lisänneen rikollisuutta (esim. myymälävarkauksia) sekä kohteena olleiden vanhempien että heidän lastensa joukossa, sekä heikentäneen lasten oppimistuloksia.

Yksi vahvimista kvasikokeellisista menetelmistä on ns. regressioepäjatkuvasasetelma. Tällainen asetelma syntyy, kun jonkin kiinnostavan riskitekijän tai "hoidon" todennäköisyys muuttuu jatkuvan muuttujan tai pistemäärän tietyssä kohdassa hyppäsmäisesti. Suomalainen esimerkki tällaisen asetelman nokkelasta hyödyntämisestä on Huttusen ja kumppaneiden tutkimus (2018), jossa tutkitaan toisen asteen koulutukseen pääsemisen vaikutusta rikollisuuteen vertaillen nuoria, joiden peruskoulun päättötodistuksen keskiarvo oli kyseisen koulutusohjelman sisäänpääsyrajan jommallakummalla puolen. Tutkimuksen mukaan kouluun pääseminen vähentää rikoksia erityisesti seurannan alkupäässä.

Toinen jo kohtalaisen paljon hyödynnetty tapa rangaistusten ankaruuden vaikutusten tutkimiseen on tuomarikohtaisen vaihtelun käyttäminen ns. instrumenttimuuttujana (Frandsen ym., 2019). Asetelman toimivuuden keskeisenä edellytyksenä ja lähtökohtana on oletus juttujen jakamisen satunnaisuudesta. Mikäli tuomareiden punitiivisuudessa

on systemaattisia eroja (=samanlaisesta rikoksesta tuomitaan erilaisia rangaistuksia), voidaan tätä vaihtelua hyödyntää kausaali-vaikutusten estimoinnissa. Tällä asetelmalla saadut tulokset rangaistusankaruuden vaikutuksista uusintarikollisuuteen, työssäkäyntiin ja myös tuomittujen lasten rikoskäyttäytymiseen ovat olleet jossain määrin ristiriitaisia (Bhuller ym., 2020; Dobbie ym., 2018).

Rekisteritutkimuksen tulevaisuudesta ja osaamistarpeista

Kriminologisen rekisteritutkimuksen kehitys näyttää seuraavan erityisesti ekonomistien³ (Angrist & Pischke, 2010) viitoittamaa tietä nostaessaan tutkimusasetelman tärkeyden valokeilaan. Kvasikokeellisten menetelmien kehitys eri tieteenaloilla (Gelman & Vehtari, 2020) onkin näin lisännyt rekisteriaineistojen tutkimuksellista arvoa, kun yhä paremmin ymmärretään, ettei mitattujen muuttujien vaihtelun ole ainoa tai edes paras tapa sulkea pois vaihtoehtoisia selityksiä aineistoissa havaitulle yhteisvaihtelulle. Rekisteriaineistoissa mitattujen muuttujien määrä on kuitenkin määritelmällisesti rajattu ja monien teoreettisesti relevanttien yksilö- ja ympäristötekijöiden mittaaminen on parhaimmillaankin puutteellista.

Luonnollisten koeasetelmien käyttö on houkuttelevaa ja kannatettavaa, mutta patenttiratkaisuksi kaikkiin vaikutustutkimuksen ongelmiin niistä ei ole. Mikä tahansa lainsäädäntömuutos ei helposti taivu tällaiseen arviointiin, koska mielekästä kontrolliryhmää ei synny muutoksen tapahtuessa kaikkialla sa-

³ Taloustieteen kentällä tehdään kasvavassa määrin rikollisuutta koskevaa tutkimusta. Taloustieteen apulaisprofessori (Texas A&M) Jen-

nifer Doleac ylläpitää kotisivuillaan listaa keskeisissä taloustieteen lehdissä julkaistuja, rikollisuutta koskevista tutkimuksista. <http://jenniferdoleac.com/resources/>

maan aikaan, tai satunnaisesti oletettu vaihtelu osoittautuukin jollain tavoin systemaattiseksi. On myös mahdollista, että uudella lainsäädännöllä tavoiteltu muutos ei syystä tai toisesta toteudu, ainakaan sellaisessa laajuudessa, että käyttäytymisvaikutusten tilastollinen arviointi olisi järkevästi mahdollista. Kotimainen esimerkki tällaisesta uudistuksesta on nuorisorangaistus, ja tuoreemman valvontarangaistuksenkin käyttö on jäänyt melko vähäiseksi (Lähtenmäki, 2019), mikä ei lupaa hyvää tutkimusasetelman kannalta. Joka tapauksessa luonnollisista koeasetelmista kiinnostuneiden tutkijoiden kannattaa varautua siihen, että rekisteriaineistojen sielunelämää ja syntyprosessia joutuu selvittämään myös laadullisin tutkimusmenetelmin. Datan yksityiskohtien penkominen voi olla puuduttavaa, mutta ”kengänpohjien kuluttaminen” (Freedman, 1991) ja tutkittavan muutoksen tosiasialliseen toteutukseen perehtyminen nostaa eittämättä tutkimuksen laatua.

Tässä kuvattujen tutkimusaiheiden tarkastelu paljastaa, että monet rekisteriaineistoilla tutkitut kysymykset ovat loppujen lopuksi aika perinteisiä, sillä esimerkiksi rangaistusten tai huono-osaisuuden vaikutuksia rikollisuuteen on pohdittu läpi tieteenalan historian. Kysymykset ovat kuitenkin edelleen ajankohtaisia, ja siksi on tärkeää, että niitä tarkastellaan tuoreilla aineistoilla ja vahvoilla asetelmilla. Vaikka tässä puheenvuorossa ovat korostuneet ne tutkimukset, jotka mielestäni selvimminkin tarjoavat uskottavaa näyttöä rikollisuuden syistä ja rangaistusten seurauksista, voidaan rekisteriaineistoja käyttää moniin muihinkin tarkoituksiin. Kuvailevamman tutkimuksen saaminen julkaistuksi voi olla vaikeampaa, mutta perustiedolle rikollisuuden kehityksestä ja rakenteesta on jatkuva tarve esimerkiksi lainvalmistelussa.

Suomalaisen ja pohjoismaisen rekisteritutkimuksen vahvuutena ja kilpailuetuna on erilaisten aineistojen laaja kirjo, ajallinen kattavuus ja niiden luotettava yhdistettävyyttä. Yksittäisten rekisteritietoa tuottavien tietojärjestelmien laatu on toinen kysymys. Keskeiset tietojärjestelmät, kuten Poliisiasiain tietojärjestelmä PATJA sekä syyttäjälaitoksessa ja käräjäoikeuksissa edelleen käytössä oleva Sakari ovat varsin iäkkäitä järjestelmiä, jotka tullaan lähivuosina korvaamaan uusilla tietojärjestelmillä. Ei siis ole syytä ajatella, että nyt jo eläkeiässä olevien tietojärjestelmien tuottaman tiedon laatu tai määrä olisivat mitenkään ainutlaatuisella tasolla kansainvälisesti vertaillen. Kokonaan toinen kysymys on, milloin uudet järjestelmät (mm. AIPA, VITJA, ROTI) otetaan käyttöön, ja millaista tietoa ne tulevat tutkijoille tarjoamaan. Siitä huolimatta, että on helppo ymmärtää tietojärjestelmäkehityksen toimivan monien vaatimusten ja toiveiden ristipaineessa, olisi toivottavaa, että datan jälkikäteen hyödyntämisen mahdollisuuksia tutkimuksessa ja hallinnossa ei unohdeta – tai otettaisi itsensänselvyytenä, johon ei erityisesti tarvitse panostaa.

Edellä totesin, että luonnollisia koeasetelmia hyödyntävien tutkijoiden kannattaa valmistautua perehtymään tutkittavan intervention tosiasialliseen toteutukseen ja sen yksityiskohtiin. Sama pätee rekisteriaineistojen käyttöön laajemminkin. On tietysti hienoa, että valmisaineistojen saatavuus on parantunut, mutta tästä huolimatta olisi hyödyllistä, jos tutkijat ymmärtäisivät perusasiat tietojärjestelmien toiminnasta, tietokannoista ja niistä prosesseista, joissa tietokantojen raakadata muokataan tilastointi- ja tutkimuskäyttöön. Aineiston esikäsittelyn (Sund ym., 2004) ja siinä tehtyjen monien valintojen merkitystä voi olla muuten vaikea hahmottaa. Näin on

erityisesti silloin, jos tutkija mieli käyttää jotain vähemmän käytettyä aineistoa, jonka erikoisuuksia ja laatu puutteita ei yleisesti tunneta. Viime kädessä raakadatan tarkkuus määrittää sen, kuinka nyansoitua analyysia sillä voi tehdä, ja parhaimmillaan tietokannan syövereiden tunteminen voi jopa herättää uusia tutkimuskysymyksiä.

Mielestäni rekisteritutkimuksesta kiinnostuneen opiskelijan kannattaa ehdottomasti panostaa modernien tilastollisten tutkimusmenetelmien opiskeluun ja kausaalipäätelyä koskevaan keskusteluun⁴. Hyviä oppikirjoja havainnoivasta aineistosta tehtävään kausaalipäätelyyn ovat muun muassa Morgan & Winship (2014), Angrist & Pischke (2009) ja Pearl (2018). Lisäksi suosittelen panostamaan aineiston käsittelytaitojen sekä datan visualisoinnin opetteluun.

FIONA-etäkäyttöjärjestelmä on tuonut rikollisuutta koskevat rekisteriaineistot aikaisempaa useamman tutkijan ulottuviin, ja siten "demokratisoinut" tutkimuksen mahdollisuuksia. Toivottavasti tämä johtaa siihen, että jatkossa yhä useampi opiskelija ja tutkija kiinnostuu rikollisuutta koskevasta rekisteritutkimuksesta. Näkisin itse, että rekisteriaineistot tarjoavat yksilöihin keskittyvän analyysin lisäksi paljon käyttämättömiä mahdollisuuksia muun muassa rangaistuskäytännön tutkimukseen sekä rikosprosessin (rikosepäilyn kulku tuomioon ja täytäntöönpanoon saakka) kokonaistarkasteluun, vaikka nykyaineistot eivät aivan optimaalisia tähän käyttöön vielä olekaan. Koneoppimismenetelmien ja esimerkiksi luonnollisen kielen käsittelyn nopea kehitys avaa myös monia uusia mahdollisuuksia rakenteettoman datan rakenteistamisen osalta. Näkisin myös mielel-

läni lisää tutkimusta rekisteriaineistojen luotettavuudesta, mahdollista harhoista ja laatu puutteista sekä niiden syistä. Tällaista tutkimusta voi pitää erityisen tärkeänä nyt, kun rekisteriaineistojen potentiaali tunnistetaan entistä laajemmin, ja tutkimuskäyttö mitä todennäköisimmin lisääntyy entisestään.

KIRJALLISUUS

- Aaltonen, M. (2013). *Socioeconomic differences in crime and victimization. A register-based study*. Helsinki: Oikeuspoliittinen tutkimuslaitos.
- Aaltonen, M., Oksanen, A., & Kivivuori, J. (2016). Debt problems and crime. *Criminology*, 54(2), 307-331.
- Allison, P. (2009). *Fixed effects regression models*. Thousand Oaks: Sage.
- Andersen, L. H., Anker, A. S. T., & Andersen, S. H. (2016). A formal decomposition of declining youth crime in Denmark. *Demographic Research*, 35, 1303-1316.
- Andersen, L. H., & Andersen, S. H. (2014). Effect of electronic monitoring on social welfare dependence. *Criminology & Public Policy*, 13(3), 349-379.
- Andersen, L. H., Dustmann, C., & Landersø, R. (2019). *Lowering welfare benefits: intended and unintended consequences for migrants and their families* (CReAM Discussion Paper Series 5/19). London: Centre for Research and Analysis of Migration.
- Andersen, S. H. (2015). Serving Time or Serving the Community? Exploiting a policy reform to assess the causal effects

⁴ "Potential outcomes" ja "Directed acyclic graphs" -lähestymistapojen yhtäläisyyksistä ja

eroista käydään vilkasta keskustelua, ks. esim. <https://arxiv.org/pdf/1907.07271.pdf>

- of community service on income, social benefit dependency and recidivism. *Journal of Quantitative Criminology*, 31(4), 537–563.
- Andersen, S. H. (2012). *Unemployment and crime: Experimental evidence of the causal effects of intensified ALMPs on crime rates among unemployed individuals*. Copenhagen: Rockwool Foundation Research Unit.
- Andersen, S. N., & Telle, K. (2019). Better out than in? The effect on recidivism of replacing incarceration with electronic monitoring in Norway. *European Journal of Criminology*. Julkaistu ennakkoon verkossa.
<https://doi.org/10.1177/1477370819887515>
- Angrist, J. D., & Pischke, J-S. (2009). *Mostly harmless econometrics. An empiricist's companion*. Princeton: Princeton University Press.
- Angrist, J. D., & Pischke, J-S. (2010). The credibility revolution in empirical economics: how better research design is taking the con out of economics. *Journal of Economic Perspectives*, 24(2), 3-30.
- Anker, A. S. T., Doleac, J. L., & Landersø, R. (2020). The effects of DNA databases on the deterrence and detection of offenders. *American Economic Journal: Applied Economics*, tulossa.
- Bennett, P., & Ouazad, A. (2020). Job displacement, unemployment, and crime: evidence from Danish microdata and reforms. *Journal of the European Economic Association*, 18(5), 2182–2220.
- Bhuller, M., Dahl, G. B., Løken, K. V., & Mogstad, M. (2020). Incarceration, recidivism, and employment. *Journal of Political Economy*, 128(4), 1269-1324.
- Blokland, A. A. J., & Nieuwebeerta, P. (2005). The effects of life circumstances on longitudinal trajectories of offending. *Criminology*, 43(4), 1203-1240.
- Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59(1), 1-12.
- D'Onofrio, B. M., Sjölander, A., Lahey, B. B., Lichtenstein, P., & Öberg, A. S. (2020). Accounting for Confounding in Observational Studies. *Annual Review of Clinical Psychology*, 16, 25–48.
- Dobbie, W., Grönqvist, H., Niknami, S., Palme, M., & Priks, M. (2018). The intergenerational effects of parental incarceration. NBER Working Paper No. 24186. Cambridge: National Bureau of Economic Research.
- Elonheimo, H., Gyllenberg, D., Huttunen, J., Ristkari, T., Sillanmäki, L., & Sourander, A. (2014). Criminal offending among males and females between ages 15 and 30 in a population-based nationwide 1981 birth cohort: results from the Finn-Crime Study. *Journal of Adolescence*, 37(8), 1269-1279.
- Fallesen, P., & Andersen, L.H. (2017). Explaining the consequences of imprisonment for union formation and dissolution in Denmark. *Journal of Policy Analysis and Management*, 36(1), 154-177.
- Fallesen, P., Geerdsen, L.P., Imai, S., & Tranæs, T. (2018). The effect of active labor market policies on crime: Incapacitation and program effects. *Labour Economics*, 52, 263-286.
- Frandsen, B. R., Lefgren, L. J., & Leslie, E. C. (2019). *Judging judge fixed effects* (NBER Working Paper No. 25528). Cambridge: National Bureau of Economic Research.
- Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology*, 21, 291-313.

- Frisell, T. (2012). *Violent crime : addressing causation with family-based methods*. Stockholm: Karolinska Institutet.
- Gelman, A., & Vehtari, A. (2020). What are the most important statistical ideas of the past 50 years? Julkaisematon käsikirjoitus. <http://www.stat.columbia.edu/~gelman/research/unpublished/stat50.pdf>
- HE 154/2012. *Hallituksen esitys eduskunnalle laeiksi tilastolain sekä maaseutuelinkeinoilastoista annetun lain 2 ja 3 §:n muuttamisesta*.
- Honkatukia, P. (2001). "Ilmoitti tulleensa raiskatuksi". *Tutkimus poliisin tietoon vuonna 1998 tulleista raiskausrikoksista*. Helsinki: Oikeuspoliittinen tutkimuslaitos ja Tilastokeskus.
- Huttunen, K., Pekkarinen, T., Uusitalo, R., & Virtanen, H. (2018). *Lost boys: access to secondary education and crime*. Helsinki: VATT.
- Kainulainen, H. (2004). *Raiskattu? Tutkimus raiskausten käsittelemisestä rikosprosessissa*. Helsinki: Oikeuspoliittinen tutkimuslaitos ja Tilastokeskus.
- Kinnunen, A. (2001). *Huumausainerikoksiin syyllistyneiden rikosura ja sosioekonomisen asema*. Helsinki: Oikeuspoliittinen tutkimuslaitos.
- Kyvsgaard, B. (2002). *The criminal career. The Danish longitudinal study*. Cambridge: Cambridge University Press.
- Landersø, R. (2015). Does incarceration length affect labor market outcomes? *The Journal of Law and Economics*, 58(1), 205-234.
- Larsen, B. Ø. (2017). Educational outcomes after serving with electronic monitoring: Results from a natural experiment. *Journal of Quantitative Criminology*, 33(1), 157-178.
- Latvala, A. (2020). Rikollisuuden syiden ja seurausten jäljillä – mahdoton tehtävä? *Haaste*, 1-2/2020.
- Lyngstad, T. H., & Skardhamar, T. (2011). Nordic register data and their untapped potential for criminological knowledge. *Crime and Justice*, 40(1), 613-645.
- Lähteenmäki, N. (2019). Valvontarangaistus jäänyt marginaaliasemaan. *Haaste*, 2/2019.
- Morgan, S. L. & Winship, C. (2014). *Counterfactuals and Causal Inference. Methods and Principles for Social Research*. Cambridge: Cambridge University Press.
- Nagin, D. (2005). *Group-based modeling of development*. Cambridge: Harvard University Press.
- Paananen, R., Ristikari, T., Merikukka, M., Rämö, A., & Gissler, M. (2012). *Lasten ja nuorten hyvinvointi Kansallinen syntymäkohortti 1987-tutkimusaineiston valossa*. Helsinki: Terveystieteiden ja hyvinvoinnin laitos.
- Pearl, J. (2018). *The book of why: the new science of cause and effect*. New York: Basic Books.
- Ramakers, A., Aaltonen, M., & Martikainen, P. (2020). A closer look at labour market status and crime among a general population sample of young men and women. *Advances in Life Course Research*. Julkaistu ennakkoon verkossa. <https://doi.org/10.1016/j.alcr.2019.100322>
- Rege, M., Skardhamar, T., Telle, K., & Votruba, M. (2019). Job displacement and crime: Evidence from Norwegian register data. *Labour Economics*, 61, 101761.
- Sariaslan, A., Larsson, H., D'Onofrio, B., Långström, N., & Lichtenstein, P. (2014). Childhood family income, adolescent violent criminality and substance misuse: Quasi-experimental total population study. *British Journal of Psychiatry*, 205(4), 286-290.

- Sariaslan, A., Larsson, H., Lichtenstein, P., & Fazel, S. (2017). Neighborhood influences on violent reoffending risk in released prisoners diagnosed with psychotic disorders. *Schizophrenia Bulletin*, 43(5), 1011–1020.
- Savolainen, J., & Sirén, R. (2013). No evidence of specific deterrence under penal moderation: imprisonment and recidivism in Finland. *Journal of Scandinavian Studies in Criminology and Crime Prevention*, 14(2), 80-97.
- Sihvonen, A-P., & Gissler, M. (2004). Rekisteritietojen käyttö sosiaalitieteellisessä tutkimuksessa 2000-luvun alussa. *Yhteiskuntapolitiikka*, 70(5), 514-520.
- Sivertsson, F., Nilsson, A., & Bäckman, O. (2019). Participation and frequency in criminal convictions across 25 successive birth cohorts: collectivity, polarization, or convergence? *Justice Quarterly*. Julkaistu ennakkoon verkossa. <https://doi.org/10.1080/07418825.2019.1699941>
- Sund, R., Nylander, O., & Palonen, T. (2004). Raa’asta rekisteriaineistosta terveystieteellisesti relevanttiin informaatioon. *Yhteiskuntapolitiikka*, 69(4), 372-379.
- Tilastokeskus (2004). *Use of Registers and Administrative Data Sources for Statistical Purposes. Best Practices of Statistics Finland* (Käsikirjoja 45). Helsinki: Tilastokeskus.