



## Kvasikokeelliset menetelmät terveydenhuollon ja terveystaloustieteen vaikutusarvioinneissa

Terveystaloustieteen yksi keskeisimmistä käsitteistä on vaikuttavuus. Vaikutuksella viitataan usein siihen, millaiseksi tutkittavana olevan kohderyhmän kiinnostuksen kohteena oleva vaste kehittyi intervention jälkeen verrattuna tilanteeseen, jossa interventiota ei olisi tehty. Lääketieteessä vaikuttavuutta tutkitaan valtaosin satunnaistettuja kokeita hyödyntäen. Satunnaistettujen kokeiden rooli on perinteisesti ollut vähäisempi yhteiskuntatieteissä, vaikka laajoja yhteiskunnallisia kokeiluita on tehty ainakin 1970-luvulta lähtien.

Terveystaloustieteen lisäarvo terveydenhuollon toiminnan järjestämisen työkaluna on nähty ensisijaisesti priorisoinnissa ja resurssien vaihtoehtoisten käyttötarkoitusten systemaattisessa tarkastelussa, ei niinkään vaikutusarvioinnin saralla. Viime vuosikymmeninä otetut menetelmälliset edistysaskeleet taloustieteessä ja niiden sovellukset terveystaloustieteessä ovat edistäneet uskottavien tutkimusasetelmien tärkeää roolia vaikutusarvioinneissa. Monien tärkeiden terveystaloustieteellisten kysymysten tapauksessa varteenotettavin vaihtoehto vaikuttavuuden arviointiin on käyttää kokeellisten asetelmien sijasta havainnoivia kvasikokeellisia asetelmia ja niihin liittyviä ekonometrisiä menetelmiä. Kvasikokeellisten menetelmien merkittävänä etuna voidaan pitää niiden soveltumista moniin tärkeisiin terveyspolitiikkaa ja terveydenhuoltoa käsitteleviin kysymyksiin ja lisäksi se, että menetelmät mahdollistavat uskottavien vaikuttavuus- ja kustannusvaikutusarviointien tekemisen vastaavasti kuin satunnaistetut koeasetelmat.

Tässä artikkelissa esitellään satunnaistetut ja keskeisimmät kvasikokeelliset tutkimusasetelmat sekä niitä hyödyntäviä suomalaisia ja kansainvälisiä terveystaloustieteen tutkimuksia. Lisäksi artikkelissa arvioidaan ns. uskottavuusvallankumouksen vaikutusta terveystaloustieteelliseen tutkimukseen. Yhtenä tavoitteena on myös havainnollistaa opiskelijoille ja tutkijoille kvasikokeellisten asetelmien hyödyntämisen mahdollisuuksia erilaisissa ajankohtaisissa sosiaali- ja terveydenhuoltoa käsittelevissä kysymyksissä.

**ASIASANAT:** Kausaalipäätely, kvasikokeellinen asetelma, luonnollinen koasetelma, terveystaloustiede, vaikuttavuus

**MIKA KORTTELAINEN, HENRI SALOKANGAS**

### YDINASIAT

- Kvasikokeelliset tutkimusasetelmat tarjoavat luotettavan mahdollisuuden arvioida sosiaali- ja terveydenhuoltoon liittyvien toimenpiteiden vaikutuksia
- Kvasikokeellisia menetelmiä hyödynnetään nykyisin laajasti taloustieteen eri osa-alueilla kuten terveystaloustieteessä

- Kvasikokeellisten menetelmien kehittämisellä ja hyödyntämisellä on ollut merkittävä vaikutus terveystaloustieteen tutkimukseen niin Suomessa kuin maailmallakin

## JOHDANTO

Terveystaloustiede on terveyden sekä sosiaali- ja terveydenhuoltojärjestelmän toiminnan tutkimista taloustieteen näkökulmasta. terveystaloustieteen piiriin kuuluvat tutkimusaiheet ovat nykyään varsin laaja-alaisia. Perinteisesti tutkimus on kohdistunut terveydenhuoltojärjestelmän toiminnan organisointiin, rahoitukseen ja palvelutuotannon järjestämiseen sekä voimavarojen ja palvelujen tehokkaaseen ja oikeudenmukaiseen kohdentumiseen. Muiden tieteenalojen piirissä erityistä huomiota terveystaloustieteen tutkimuksen saralla on perinteisesti saanut taloudellinen arviointi. terveystaloustieteen lisäarvo terveydenhuollon toiminnan järjestämisen työkaluna on nähty ensisijaisesti juuri priorisoinnissa ja resurssien vaihtoehtojen käyttötarkoitusten systemaattisessa tarkastelussa, ei niinkään vaikutusarvioinnin saralla. Tästä on kenties lähtöisin ajoittain esille tullut virheellinen ajatus, että taloustieteessä ollaan erityisen kiinnostuneita kustannuksista tai kustannusten minimoimisesta. Tästä terveystaloustieteessä ei kuitenkaan ole kyse. Nykytutkimussuuntausta terveystaloustieteessä enemmänkin kuvaa pyrkimys hyödyntää tutkimusasetelmia, joiden pohjalta voidaan tehdä luotettavia vaikuttavuus- ja kustannusvaikutavuusarviointoja.

Intervention, altistuksen tai toimenpiteen vaikutuksella viitataan yleensä siihen, millaiseksi tutkittavana olevan kohderyhmän kiinnostuksen kohteena oleva vaste kehittyi intervention jälkeen verrattuna tilanteeseen, jossa interventiota ei olisi tehty. Lääketieteessä vaikuttavuutta pyritään yleensä tutkimaan satunnaiskokeita hyödyntäen. Myös terveystaloustieteessä satunnaiskokeita on hyödynnetty jonkin verran, mutta niiden käyttöön liittyy monissa sovelluksissa rajoitteita. Esimerkiksi eettisistä syistä johtuen jotain terveystaloustieteen toimenpidettä tai uudistusta ei ole mahdollista satunnaistaa vain osalle ihmisistä. Koska vaikutusten arviointi on kuitenkin erittäin keskeisessä roolissa terveystaloustieteessä, tarvitaan mahdollisimman luotettavaa tutkimusnäyttöä havainnollisista tai ei-satunnaisista asetelmista. Viime vuosikymmeninä otetut menetelmälliset edistysaskeleet taloustieteessä ja niiden sovellukset terveystaloustieteessä ovat edistäneet uskottavien tutkimusasetelmien tärkeää roolia vaikutusarvioinneissa.

Monien tärkeiden terveystaloustieteellisten kysymysten tapauksessa varteenotettavin vaihto-

ehto vaikuttavuuden arviointiin on käyttää koekokeellisten asetelmien sijasta havainnollisia kvasikoekokeellisia asetelmia ja niihin liittyviä ekonometrisiä menetelmiä. Kvasikoekokeellisten menetelmien merkittävänä etuna voidaan pitää niiden soveltumista moniin tärkeisiin terveystaloustieteen ja terveydenhuollon käsitteleviin kysymyksiin ja lisäksi se, että menetelmät mahdollistavat uskottavien vaikuttavuus- ja kustannusvaikutavuusarviointien tekemisen vastaavasti kuin satunnaistetut koekokeet.

Vuonna 2021 taloustieteen Nobel-palkinnon saaneiden David Cardin, Joshua Angristin ja Guido Imbensin empiirinen ja metodologinen tutkimus kvasikoekokeellisista menetelmistä on ollut tärkeässä roolissa menetelmien laajemman hyödyntämisen kannalta. Palkittujen työllä on ollut myös laaja vaikutus taloustieteen eri osa-alueiden kuten terveystaloustieteen tutkimukseen ja opetukseen sekä heijastumavaikutuksia myös muihin tieteenaloihin (1, 2). Tärkeisiin kausaalikysymyksiin vastaaminen on myös lisännyt taloustieteellisen tutkimuksen yhteiskunnallista merkitystä.

Tässä artikkelissa esittelemme satunnaistetut ja keskeisimmät kvasikoekokeelliset tutkimusasetelmat sekä niiden sovelluksia terveystaloustieteellisissä ja terveydenhuoltoon liittyvissä tutkimuksissa. Artikkelin ei ole kirjallisuuskatsaus kvasikoekokeellisiin menetelmiin, vaan tavoitteena on ennen kaikkea tarjota johdanto aiheeseen sekä havainnollistaa kvasikoekokeellisten asetelmien hyödyntämisen mahdollisuuksia erilaisissa ajankohtaisissa sosiaali- ja terveydenhuollon käsittelevissä kysymyksissä. Lisäksi arvioimme menetelmien kehittämisen ja laajemman hyödyntämisen vaikutuksia terveystaloustieteelliseen tutkimukseen ja opetukseen.

## SATUNNAISTETUT KOEASETELMAT

Lääketieteessä vaikuttavuutta pyritään tutkimaan kontrolloituja satunnaiskokeita (randomized control trials, RCTs) hyödyntäen. Satunnaiskokeita on tavanomaisesti pidetty empiirisen tutkimuksen kruununjalokivenä (engl. ”gold standard”). Yksinkertaisessa satunnaiskokeen versiossa tutkimusjoukko jaetaan kahteen osaan: koe-ryhmään, johon kohdistuu interventio, ja verrokkiryhmään, johon ei kohdistu interventiota. Tämän tutkimusmenetelmän keskeinen etu on se, että satunnaiskokeen verrokkiryhmä on satun-

naistamisen myötä niin havaituilta kuin havaitsemattomilta taustatekijöiltään keskimäärin samanlainen koeryhmän kanssa erityisesti suurissa tutkimusryhmissä. Jos satunnaiskoe toteutetaan asianmukaisesti, erot kiinnostuksen kohteessa olevassa vasteessa tutkimusryhmien välillä voidaan uskottavasti osoittaa johtuvan toteutetusta interventtiosta (virhemarginaali huomioonottaen). Vasteen keskiarvo verrokkiryhmässä edustaa tällöin intervention kontrafaktuaalia eli tilannetta, jossa koeryhmään ei olisikaan kohdistettu interventiota (3). Satunnaiskokeen tulokset ovat ns. sisäisesti valideja, ts. kertovat luotettavasti intervention vaikutuksista tarkastellun tutkimusryhmän puitteissa.

Hyvin toteutetussa satunnaiskokeessa koe- ja verrokkiryhmän erotus keskimääräisessä vasteen arvossa on toimenpiteen keskimääräisen vaikutuksen (average treatment effect) harhaton estimaatti tutkittavassa populaatiossa. Toisin kuin havaintoaineistoon perustuvissa tutkimuksissa, satunnaiskokeen puitteissa kausaalipäätelyyn ei tarvitse mallintaa eri muuttujien keskinäisriippuvuuksia, eikä kiinnittää erityistä huolta taustamuuttujien kontrolloimiseen tai sekoittaviin tekijöihin, jotka vaikuttaisivat samanaikaisesti kiinnostuksen kohteena olevaan vasteeseen ja toimenpiteen saamisen todennäköisyyteen. Havaintoaineistoon perustuvat vertailut kärsivät usein ns. valikoitumisharhasta eli koeryhmään saattaa päätyä ihmisiä, jotka pärjäisivät ilman hoitotoimenpidettä (ts. interventiota) verrokkiryhmään verrattuna suhteellisen hyvin (positiivisesti valikoitunut) tai suhteellisen huonosti (negatiivisesti valikoitunut). Tällöin verrokkiryhmä ei ole uskottava kontrafaktuaali eli ei kerro luotettavasti mitä koeryhmälle olisi keskimäärin tapahtunut ilman hoitotoimenpidettä. Tämän kaltaiseen päätelmään päätyivät Sacks et al. (4) vertaillaan kuuden eri hoitotoimenpiteeseen liittyvien 50 satunnaistettuun kontrolloituun kokeeseen perustuvien tutkimusten tuloksia 56 havaintoaineistoihin perustuvien tutkimusten tuloksiin. Yli kolme neljänestä havaintoaineistoihin perustuvista tutkimuksista arvioi hoitojen olevan tehokkaita, siinä missä vain 20 prosenttia RCT-tuloksista antoi samansuuntaisia tuloksia. Tämänkaltaiset tutkimukset ovat osaltaan kasvattaneet näkemystä satunnaistettujen kontrolloitujen kokeiden näytön ylivertaisesta luotettavuudesta suhteessa muihin tutkimusmenetelmiin,

vaikka tarkasti suunniteltujen ja rajattujen havaintoaineistoihin perustuvien tutkimusten tulokset eivät välttämättä paljonkaan eroaisi satunnaiskokeiden vastaavista (5).

Vaikka satunnaistaminen poistaa hoitotoimenpiteeseen liittyvän valikoitumisen tutkimuspopulaation sisällä, satunnaiskokeeseen liittyy kuitenkin eräitä tärkeitä haasteita. Huomionarvoisin haaste liittyy tulosten ns. ulkoiseen validiteettiin eli edustavuuteen (engl. external validity). Ulkoinen validiteetti on yleisin kliinikoiden esittämä kritiikki satunnaiskokeita kohtaan. Vaikka kritiikillä vaikuttaa olevan perusteensa, ulkoista validiteettia ei ole erityisemmin otettu huomioon lääketeollisuudessa tai lääketieteen tiedejulkaisuissa (6). Ulkoista validiteettia voi uhata mm. tutkimusryhmän rekrytointiprosessi ja -kriteerit, alue, hättävaiikutusten puutteellinen raportointi, hoidon keskeyttäminen, osallistumiskato, tai seuranta-aika. Yksittäisiin satunnaiskokeisiin liittyviä tuloksia saatetaan tulkita varsin rohkeasti yleistettävyyden sekä vaikutusten aikahorisontin osalta (7).

Ensimmäisen kerran satunnaiskokeisiin viitattiin "gold standard" -termillä Alvan Feinsteinin ja Ralph Horwitzin *The New England Journal of Medicine* (NEJM) -lehden artikkelissa vuodelta 1982 (8). Vaikka termiä on sittemmin käytetty kuvaamaan monessa yhteydessä satunnaiskokeiden erityisasemaa luotettavan kausaalipäätelmien lähteenä, Feinstein ja Horwitz (9) itse asiassa kannustivat hyödyntämään muita tutkimusasetelmia ja pitämään satunnaistettuja kokeita enemmänkin teoreettisena ideaalina kuin välttämättömänä tutkimusasetelmana tieteen kehityksessä.

Terveystaloustieteessä kokeellisia tutkimusasetelmia on käytetty jo 1970-luvulta lähtien, mutta toistaiseksi niitä on pystytty hyödyntämään vain tiettyihin tai hyvin rajattuun joukkoon tutkimuskysymyksiä. Kuuluisin esimerkki lienee vuosien 1974 ja 1981 välissä Yhdysvalloissa toteutettu RAND Health Insurance Experiment -satunnaiskoe, joka käsiteli terveysvakuutusten korvaustasojen vaikutuksia terveyspalvelujen käyttöön ja kustannuksiin (10). Toinen tuoreempi esimerkki on Oregonissa vuonna 2008 toteutettu Medicaid-vakuutuksen laajennusta koskeva satunnaiskoe (11).

Myös Suomessa satunnaistettuja koeasetelmia on hyödynnetty varsin laajasti erilaisissa lää-

ketieteellisissä sovelluksissa, mutta toistaiseksi hyvin vähän varsinaisissa terveystaloustieteellisissä kysymyksissä. Satunnaistamista on kuitenkin pystytty hyödyntämään mm. informaatiokokeilujen tapauksessa. Esimerkki tällaisesta informaatio-ohjaukseen perustuvasta tutkimuksesta on Suomessa toteutettu klusteri-satunnaistettukoe, jossa lähetettiin muistutuskirjeitä influenssarokotteisiin liittyen yli 65-vuotiaalle henkilölle kahdella eri alueella (12). Tutkimuksessa havaittiin, että informaatiokirjeillä oli positiivinen vaikutus rokottautumiseen erityisesti niiden keskuudessa, jotka eivät olleet ottaneet aiemmin influenssarokotetta. Tutkimus on esimerkki kustannustehokkaasta vaikuttamisesta terveyskäyttäytymiseen, sillä kirjeiden lähettäminen on varsin edullista ja vaikuttavuuden seuranta onnistuu suomalaisista rekistereistä saatavin tiedoin suhteellisen vähäisin kustannuksin.

### **KVASIKOKEELLISET MENETELMÄT**

Vaikka kontrolloituja satunnaiskokeita voidaan pitää yleisesti ideaalina asetelmana, on usein hyvin haastavaa ja toisinaan jopa mahdotonta hyödyntää satunnaistettuja koeasetelmia taloustieteen ja myös sosiaali- ja terveydenhuollon eri toimenpiteiden vaikuttavuuden arvioinnissa. Tähän ajatukseen ovat perustuneet viime vuosikymmeninä otetut taloustieteen menetelmälliset edistysaskeleet havainnollisiin asetelmiin perustuvissa vaikutusarvioinneissa sekä niiden sovellukset terveystaloustieteessä (1). Taloustieteessä on vahvistunut pyrkimys löytää havainnollisia vertailuasetelmia, joissa hoitotoimenpiteen, politiikkatoimen tai muun tapahtuman toteutumiseen liittyy sattuman kaltaista vaihtelua tietyssä tutkimusryhmässä. Kvasikokeellisilla asetelmilla (quasi-experimental designs) tarkoitetaan asetelmia, joissa yksilöiden (tai muiden havaintoyksikköjen) valikoituminen koe- ja verrokkiryhmään määräytyy mekanismeilla, joka tutkittavien yksilöiden näkökulmasta muistuttaa satunnaistamista. Tähän termiin liittyy läheisesti luonnollinen koeasetelma (natural experiment), joka yleensä viittaa luonnostaan tapahtuviin äkillisiin muutoksiin elinympäristössä ja niiden sattumankaltaiseen altistukseen perustuviin tutkimusasetelmiin, jossa ihminen ei vaikuta koeasetelman muodostamiseen. Tässä artikkelissa keskitymme kvasikokeellisiin asetelmiin, joissa vertailuryhmät ovat viime kädessä tutkijoiden valitseamia.

Kvasikokeellisten vertailuasetelmien ideana on hyödyntää satunnaisen kaltaista vaihtelua tutkittavassa toimenpiteessä tai tapahtumassa. Toimenpide ei useinkaan ole satunnainen tutkimusryhmässä, vaan satunnaisen kaltainen tietyissä rajatuissa vertailuryhmissä tietyt taustatekijät mahdollisesti vakioiden (conditional mean independence). Usein kvasikokeellinen asetelma muodostuu jonkin politiikkamuutoksen tai intervention seurauksena, mutta myös erilaiset sattumat (esim. tarkka syntymäajankohta) ja yhteiskunnan säännöt voivat luoda tutkimusasetelmia, jossa osa henkilöistä tai vertailuysiköistä joutuu tietyn toimenpiteen tai käsittelyn kohteeksi satunnaisen kaltaisesti. Tämäntyyppisten asetelmien keskeinen etu on se, että ne voivat vähentää merkittävästi systemaattisia, toimenpiteestä riippumattomia, eroja koe- ja verrokkiryhmän välillä. Esimerkiksi syntymäaikaan perustuva rajaus politiikkatoimenpiteessä voi satunnaistamisen lailla parhaimmillaan tehdä vertailtavista ryhmistä keskimäärin hyvin samankaltaiset.

Mikäli koe- ja vertailuryhmä eroavat havaituilta ominaisuuksilta toisistaan, huolenaiheena on se, että erot vasteessa ryhmien välillä muodostuvat altistuksen lisäksi sekoitettavien tekijöiden johdosta. Mikäli on syytä olettaa, että altistukseen valikoituminen tapahtuu yksinomaan havaittujen ominaisuuksien perusteella, on mahdollista vähentää oleellisesti sekoittavien tekijöiden luomaa harhaa kontrolloimalla regressiomallissa havaittuja, ennen altistetta mitattuja ominaisuuksia tai toteuttaa tilastollinen kaltaistus ryhmien välillä havaittuihin ominaisuuksiin perustuen. Mikäli on kuitenkin syytä olettaa, että altistukseen valikoituminen perustuu oleellisesti havaitsemattomiin tekijöihin (esim. geenit, terveyskäyttäytyminen), niin kaltaistumenetelmät tai havaittujen tekijöiden kontrollointi eivät auta kausaalivaikutuksen identifioinnissa (13).

Kvasikokeellisia asetelmia on hyödynnetty epidemiologiassa ja psykologiassa jo pitkälti ennen 1980-lukua (14–16), mutta niin taloustieteessä kuin muillakin tieteenaloilla niiden rooli empiirisessä tutkimuksessa on ollut käytännössä varsin vähäinen ennen 2000-lukua. Tähän on ollut osittain syynä menetelmien asteittainen kehittyminen, mutta tärkeä tekijä on ollut myös uskottavien empiiristen sovellusten yleistyminen vähitellen 1980-luvulta lähtien.

Seuraavassa esittelemme tarkemmin kolmea keskeisintä ja tunnetuinta kvasikokeellista menetelmää, (i) erotusten erotus -menetelmää, (ii) regressioepäjatkuvuus- ja (iii) instrumenttimuuttujamenetelmää. Lisäksi esittelemme menetelmien sovellutuksia terveystaloustieteen piirissä. Keskitämme kyseisiin menetelmiin, koska ne ovat viime vuosikymmeninä nousseet suosituiksi nimenomaan soveltavissa terveystaloustieteen tutkimuksissa. Tästä hyvänä havainnollistuksena on näitä menetelmiä hyödyntävien artikkeleiden osuukien muutokset vuosina 2009–2022 alan johtavassa Journal of Health Economics -lehdessä. Taulukon 1 mukaan erityisesti erotusten erotus -menetelmää ja regressioepäjatkuvuusmenetelmää hyödyntävien tutkimusten määrä on nousut merkittävästi viimeisen 10 vuoden aikana suhteessa esimerkiksi kaltaistamismenetelmiin.

Taulukko 1. Kvasikokeellisten menetelmien yleisyys Journal of Health economics -lehdessä vuosina 2009–2022

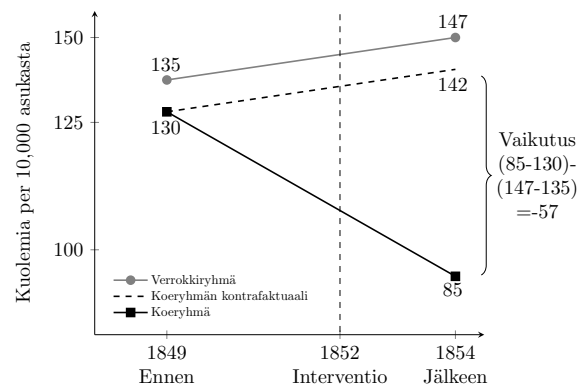
	Regressio- epä- jatkuvuus	Instru- mentti- muuttuja	Erotusten erotus	Kaltaistus
2009–2010	2,3	3,5	2,9	2,9
2011–2013	2,6	5,1	7,2	6
2014–2016	2,7	3,9	6,2	3,5
2017–2019	3,8	3,1	6,5	3,8
2020–2022	5,2	2,1	10,4	3,5

Seloste. Laskelmat perustuvat IDEAS/RePEc -tietokannassa tehtyihin hakuihin termein “Regression discontinuity”, “Difference-in-differences”, “Instrumental variable” ja “matching” Journal of Health Economics -lehdessä. Julkaisumäärät suhteutettiin ajanjaksolla julkaistujen artikkeleiden kokonaismääriin ajanjaksolta 2009–2022. Osa tutkimusartikkeleista hyödyntää useampaa kuin yhtä mainituista menetelmistä.

## EROTUSTEN EROTUS -MENETELMÄ

Erotusten erotus (difference-in-differences, DiD) -menetelmässä verrataan koe- ja verrokkiryhmän välisiä eroja vasteessa ennen ja jälkeen tietyn tapahtuman. Menetelmän uskottavuus vaikutusarvioinnissa perustuu siihen oletukseen, että koeryhmän kokema muutos vasteessa olisi samansuuntainen verrokkiryhmän kanssa siinä hypoteettisessa tilanteessa, että koeryhmään ei olisi kohdistunut interventiota. Tällöin verrokkiryhmän kehitys vasteessa muodostaa koeryhmälle uskottavan kontrafaktuaalin.

Erotusten erotus -menetelmään on luontevaa tutustua perehtymällä sen ensimmäiseen ja ehkäpä tunnetuimpaan esimerkkiin, John Snow’n koleratutkimukseen 1800-luvun puolivälissä. Tuohon aikaan Thamesjoki Lontoossa oli pahasti saastunut sinne ohjattujen likaviemäriverien johdosta, mutta Thamesin alajuoksulta otetun juomaveden yhteyttä koleran leviämiseen ei oltu vielä todistettu. Vallitseva käsitys koleran leviämisestä perustui myrkyllisiin höyryihin, mutta John Snow uskoi taudinaiheuttajien leviävän ulostevesien sekoittumisesta puhtaaseen veteen. Osoittaakseen teoriansa paikkansa pitävyyden hän tarkasteli kuolinlukuja ja vertasi keskenään Lontoon suurkaupungin alueita, joista osa sai vetensä pääosin vesiputkikyhtiöltä (Lambeth) jonka vesi tuli Thamesjoen yläjuoksulta (koeryhmä) ja osa kokonaan Thamesjoen alajuoksulta Southwark-Vauxhall-putkikyhtiöltä (verrokkiryhmä). Otolliseksi vertailun teki se, että Lambeth-putkikyhtiö siirsi vedenottonsa seurannan mitausajankohtien (1849 ja 1854) välissä vuonna 1852 Thamesin alajuoksulta yläjuoksulle paikkaan, jossa Thamesin vesi oli vielä puhdasta. Koska erot näiden alueiden välillä asuinväestön ammattitaustassa sekä demografisissa taustatekijöissä olivat varsin yhtäläiset, John Snow pystyi uskottavasti argumentoimaan alueiden välisten erojen muutoksen johtuvan vedenotto paikasta. Kuvio 1 demonstroi tätä päättelyä perinteisen erotusten erotus -vertailun puitteissa.



Kuvio 1. Erotusten erotus -menetelmän havainnollistus John Snow’n koleratutkimukseen perustuen.

Seloste. Kuolleisuusasteet perustuvat John Snow’n tilastoisiin kuolemiin ja väestötietoihin (13). Snow’n tutkimuksesta voidaan toteuttaa erilaisia alueellisia vertailuja. Hyödynnämme esimerkissämme vertailua, jotka on taulukoitu Nick Huntington-Kleinin (17) kirjassa The Effect: An Introduction to Research Design and Causality.

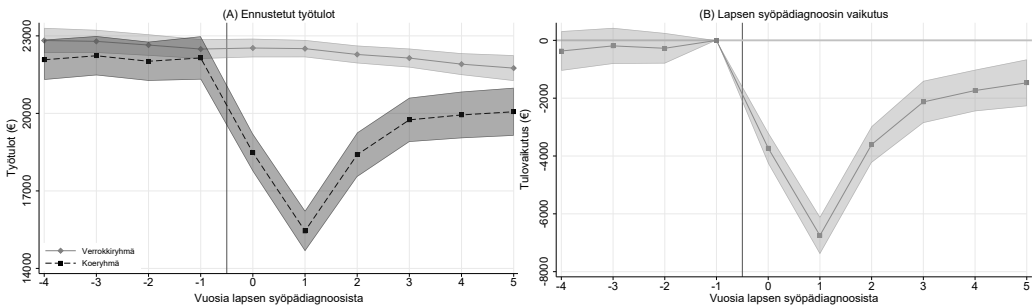
Snow'n tutkimuksessa kuolemat suhteessa väestöön olivat koe- ja verrokkiryhmässä melko lailla samanlaiset vuonna 1849 eli kolme vuotta ennen interventiota, ts. Lambeth-yhtiön vedenottopaikan muuttamista. Olettaen, että koe- ja verrokkialueisiin ei kohdistunut vuosien 1849 ja 1854 välillä muita muutoksia, jotka voisivat vaikuttaa eriäviin kuolinasteisiin, voidaan olettaa, että koe- ja verrokkiryhmän kuolinasteet etenivät samansuuntaisten trendien mukaisesti, jos koeryhmän vedenottopaikka olisi ollut edelleen Thamesin alajuoksulla (kontrafaktuaali katkoviivana). Tällöin vedenottopaikan vaihdoksen vaikutus kuolinasteeseen on koeryhmän toteutunut intervention jälkeen havaittu kuolinaste vähennettynä yhtenevien trendien oletusten mukaisesti kontrafaktuaalin kuolinasteesta.

Erotusten erotus -menetelmä on perusperiaatteiltaan samankaltainen satunnaistetun kontrolloidun koeasetelman kanssa sillä erotuksella, että päättely perustuu koe- ja verrokkiryhmän vasteen kehitystrendien samankaltaisuuteen, ei niinkään taustatekijöiden osalta (keskimääräisesti) identtisiin ryhmiin. Tätä yhtenevien trendien oletuksen (parallel trends assumption) toteutumista

pidetään taloustieteessä kriittisenä uskottaville vaikutusarvioinneille erotusten erotus -menetelmää hyödyntävissä tutkimuksissa.

Koe- ja verrokkiryhmien mahdolliset erot havaituissa tai havaitsemattomissa tekijöissä voivat kuitenkin haastaa yhtenevien trendien oletuksen uskottavuutta. Osittain tästä syystä empiirisessä tutkimuksessa on siirrytty yhä enemmän hyödyntämään yleisempää, useamman kuin kahden periodin erotusten erotus -mallia, joissa sallitaan vaikutusten muuttuvan ajassa. Näitä kutsutaan dynaamisiksi erotusten erotus -malleiksi tai event study -malleiksi.

Havainnollistamme seuraavaksi dynaamista erotusten erotus -menetelmää sekä keskeistä yhtenevien trendien oletusta tarkastelemalla lapsen syövän vaikutusta äidin työssäkäyntiin suomalaisen rekisteritutkimukseen (18) perustuen (Kuvio 2). Kyseisessä tutkimuksessa tarkastellaan vanhempien tulokehitystä ennen ja jälkeen lapsen syöpädiagnoosin (koeryhmä) ja verrataan tätä koulutuksen, syntymävuoden ja perheeseen suhteen kaltaistettuihin vanhempiin/perheisiin (verrokkiryhmä).



Kuvio 2. Tutkimus lapsen syövän vaikutuksesta äidin tulokehitykseen erotusten erotus -menetelmää hyödyntäen.

Seloste. Äidin keskimääräisten tulojen kehitys vuosina ennen (aikaperiodit -4 - -1) ja jälkeen (aikaperiodit 1-5) lapsen 1. sairaalakäynnin yhteydessä todettuun syöpädiagnoosin (aikaperiodi 0) suhteessa kaltaistettuihin äiteihin. Paneelissa A visualisoidaan tulotasojen (vuoden 2019 hinnoissa) keskimääräinen kehitys syöpäperheiden äideille ja heidän kaltaistetuille verrokeilleen. Paneelissa B demonstroidaan syövän vaikutusta äidin tuloihin vertaamalla syöpää edeltävän vuoden erotusta (referenssipiste) syöpäperheiden ja verrokkiperheiden äitien keskimääräisessä tulotasossa sitä seuraavien vuosien tulotason erotuksiin. Väritetty alue kuvaa 95 prosentin luottamusväliä. Kaltaistetut äidit ovat syntymävuoden, perherakenteen ja koulutustason osalta samankaltaisia syöpäperheiden äitien kanssa. Kaltaistetuille perheille syöpädiagnosi voidaan tulkita ns. plasebo-diagnosina.

Kuvio 2 havainnollistaa tätä tutkimusasetelmaa. Kuviossa 2A on piirretty ns. dynaamisen erotusten erotus -mallin rekisteriaineistoon perustuvat ennusteet äidin tuloista ennen ja jälkeen lapsen syöpädiagnoosin (periodi 0). Kuvioista nähdään, että tulotasot ovat hyvin samanlaisia ja kehittyvät samansuuntaisesti ennen syöpädiagnoosia koe- ja verrokkiryhmässä yhtenevien trendin oletuksen mukaisesti. Syöpädiagnoosivuotena äidin tulot laskevat noin 3500 euroa ja sitä seuraavana vuonna jopa 6000 euroa suhteessa syöpää edeltävään tasoon. Vaikutusarvioinnin kannalta oleellista on kuitenkin verrata syöpäperheen äidin tulotason kehitystä suhteessa verrokkiryhmän äiteihin, joiden lapsi ei sairastunut syöpään. Paneelissa B tarkastellaan erotusten erotus -mallin mukaisesti koe- ja verrokkiryhmän äitien keskimääräisten tulojen erotusta suhteessa syöpädiagnoosia edeltävän vuoden erotukseen. Aikaperiodien 0–5 estimaatit kuvastavat lapsen syövän vaikutusta äidin tulotasoon.

Tämä tutkimus kuuluu viime vuosina yleistyneeseen ns. terveysoskoi-kirjallisuuteen (esim. 19–21), jossa tarkastellaan terveydentilan äkillistä heikkenemistä tai terveydentilan heikkenemiseen liittyvää yllätyksellisen tiedon vaikutusta työssäkäyntiin ja työtuloihin. Tutkimuksessa tarkasteltu syöpädiagnoosi ei ole suinkaan satunnainen, mutta syöpään sairastuneiden ja samankäisten verrokkien taustatekijät eivät tilastollisesti merkitsevästi eroa toisistaan kun vanhempien ikä, koulutus ja perherakenne kaltaistetaan. Lapsen syöpä on esimerkki terveystapahtumasta, joka on ehdollisesti satunnaisen kaltainen ja tällöin kaltaistaminen on hyödyllinen työkalu kausaalitulkintojen tekemisessä.

Käytetyssä esimerkissä tulojen erotukset ovat tilastollisesti ei-merkitseviä syöpää edeltävinä vuosina (aikaperiodit (-2 – -4) vertailupisteeseen (aikaperiodi -1) nähden, mikä antaa vahvaa tukea yhtenevien trendien oletuksen toteutumiselle ennen syöpädiagnoosia. Tämän lisäksi on hyvä verrata taustatekijöitä koe- ja verrokkiryhmän välillä ennen terveystapahtumaa kuten on tapana satunnaiskokeen tapauksessa. Tässä nimenomaisessa tutkimuksessa havaitaan, että lapsen syöpä ei ole systemaattisesti yhteydessä vanhempien tulotasoon, psykiatriseen historiaan tai koulutustasoon, mikä entisestään tukee vertailuryhmän uskottavuutta koeryhmän kontrafaktuaalina myös pidemmän aikavälin vaikutusarvioinneissa.

Erotusten erotus -menetelmää on viime vuosien aikana hyödynnetty myös monissa muissa suomalaisissa terveystaloustieteellisissä tutkimuksissa. Menetelmän avulla on tutkittu mm. lääkemääräyspalautteen vaikutuksia opioidien määräämiseen (22); sähköisten reseptien käyttöönoton vaikutuksia lääkkeiden käyttöön (23), lääkkeiden määrääjien keskinäiseen koordinaatioon ja määräämisen laatuun (24); terveydenhuollon valinnanvapausuudistuksen vaikutuksia (25), lääkkeiden hintasääntelyuudistusten vaikutuksia lääkemenoihin (26) sekä hoitajakäyntien asiakasmaksun vaiheittaisen käyttöönoton vaikutuksia terveyskeskuksen hoitaja- ja lääkärikäyntien määrään (27). Suurin osa näistä tutkimuksista perustuu vaiheittain eri alueilla toteutettuihin politiikkamuutoksiin.

Näissä tutkimuksissa hyödynnetään tyypillisesti alueiden välistä ajallista vaihtelua uudistusten käyttöönotossa. Tällöin kontrafaktuaali reformien läpikäyneille alueille muodostuu alueista, jotka läpikäyvät saman reformin mutta joitain vuosia myöhemmin. Vaiheittain toteutettuihin reformeihin liittyy kuitenkin huolia (13,28), jotka näissä esimerkeissä on otettu huomioon hyödyntämällä viimeaikaista kehitysaskelia soveltavassa ekonometriassa (29–31).

Aluetason interventioissa koe- ja vertailuryhmät saattavat toisinaan poiketa toisistaan tilastollisesti merkitsevästi. Tämä ei kuitenkaan estä kausaalitulkintojen tekemistä erotusten erotus -menetelmässä. Oleellisinta on se, että reformin ajoitus tulee uskottavien vaikutusarviointien näkökulmasta olla satunnaisen kaltainen vasteen potentiaalisten kehityskulkujen osalta, ei niinkään alueiden taustatekijöiden osalta.

Mainittakoon vielä, että erotusten erotus -menetelmään liittyy läheisesti parin viimeisen vuosikymmenen aikana kehitetty synteettinen kontrollimenetelmä. Synteettinen kontrollimenetelmä on perustavilta taustaoletuksiltaan erotusten erotus -menetelmän kaltainen (32–33), mutta sen hyödyt pääsevät esille lähinnä aggregoidussa paneeliaineistossa, jossa tietty alueeseen kohdistuu interventio (esimerkiksi lakiuudistus) muiden alueiden toimiessa potentiaalisina verrokkialueina. Tässä menetelmässä hyödynnetään algoritmia, joka muodostaa useista alueista aluepainojen avulla yhden ns. synteettisen alueen, jonka tehtävänä on toimia uskottavana kontrafaktuaalina interventioalueelle. Huomionarvoi-

nen esimerkki synteettisen kontrollimenetelmän hyödyntämisestä on tutkimus koronaepidemian aikaisten maskipakkojen vaikutuksista infektioiden leviämiseen Saksassa (34). Suomessa synteettistä kontrollimenetelmään on käytetty terveystaloustieteen piirissä hyvin vähän. Eräässä suomalaisessa sovelluksessa arvioitiin Kainuun hallintokokeilun kustannusvaikutuksia sosiaali- ja terveystaloustieteen osalta (35).

## REGRESSIOEPÄJATKUVUUSMENETELMÄ

Toinen taloustieteen eri sovellusalueilla ja myös terveystaloustieteessä viime vuosina yleistynyt kvasikokeellinen menetelmä on niin sanottu regressioepäjatkuvuusasetelma (regression discontinuity design, RD). Kyseisen tutkimusasetelman ideana on hyödyntää jonkin säännön seurauksena syntyneitä asetelmaa, missä säännön raja-arvo (esim. yliopistoon pääsyn pisteraja) jakaa havaintoyksiköt lähes satunnaisesti kahteen ryhmään, jotka ovat vertailukelpoisia yksiköiden muiden havaittujen ja ei-havaittujen ominaisuuksien suhteen rajaarvonn kohdalla, mutta aiheuttaa epäjatkuvuuden mielenkiinnon kohteena olevassa muuttujassa (esim. myöhempi terveydenhuoltopalvelujen käyttäminen (36)). Yleensä päätössääntö, kuten henkilön ikä tai syntymäaika, määrittää sen kenelle toimenpide kohdistuu sekä ketkä kuuluvat koe- ja vertailuryhmään. Käytännössä päätössääntö aiheuttaa epäjatkuvuuden toimenpiteen saamisen todennäköisyyteen, mistä myös menetelmän nimi juontuu. Menetelmän ideana on verrata juuri ja juuri tai rimaa hipoen toimenpiteen saaneita niihin, jotka jäivät rimaa hipoen toimenpidettä vaille.

Terveystaloustieteessä eräs klassinen esimerkki satunnaisen kaltaisesta hoidon epäjatkuvuudesta liittyy pikkukeskosien määrittämiseen. Suomessa pikkukeskoseksi kutsutaan kaikkia lapsia, jotka ovat syntyneet ennen raskausviikkoa 32 tai niitä lapsia jotka ovat syntyneet 32 raskausviikolla tai sen jälkeen alle 1500 gramman painoisina (sisältäen 1500g). Tässä esimerkissä menetelmä hyödyntää syntymäpainoon liittyvää katkoskohtaa vastasyntyneiden sairaalahoidossa.

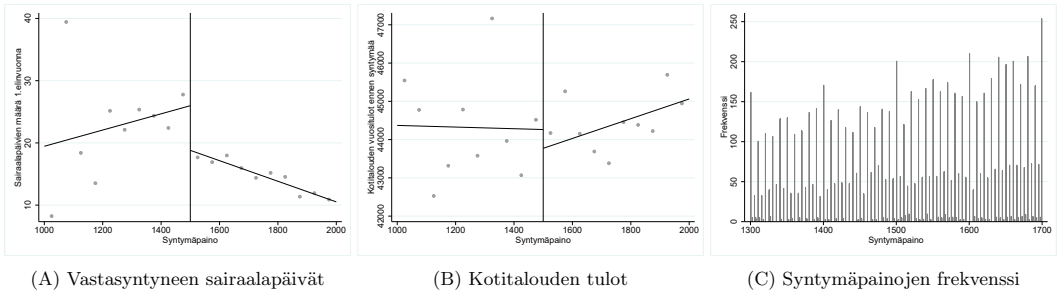
Havaintoaineistoista tehdyt tavanomaiset vertailut pikkukeskosten ja muiden lasten välillä eivät kuitenkaan johda luotettaviin pikkukeskoshoidon vaikuttavuusarvioihin. Lääketieteen kirjallisuudessa on laajasti dokumentoitu alhaisen

syntymäpainon negatiivinen yhteys terveydelisiin ja kognitiivisiin voimavaroihin. Yleisesti ottaen pikkukeskoset eroavat painavammista lapsista syntymäterveyden osalta, ja pikkukeskoshoitoihin valikoituu lapsia, jotka suoriutuisivat koulussa ja työelämässä heikommin, ja joiden terveydentila olisi heikompi suhteessa normaalipainoisiin myös siinä tapauksessa, että he jäisivät ilman pikkukeskoshoittoa. Jos kuitenkin tarkastelua rajoitettaisiin raskausviikolla 32+, ja aivan 1500 gramman rajan tuntumassa syntyneisiin, valikoituvuus pikkukeskoshoittoon pienenee käytännössä olemattomiin. Toisin sanoen lasten potentiaaliset elämänkulut olisivat odotusarvoisesti samankaltaisia 1499 gramman ja 1501 gramman painoisten lasten välillä. Nämä lapset ovat siis omien synnynnäisten voimavarojensa sekä vanhempien taustatekijöiden (esim. koulutus ja tulotaso) osalta keskimäärin samankaltaisia, mutta heitä erottaa vain pikkukeskosuus-määrittelmä. Alle 1500 gramman painoiset lapset luokitellaan pikkukeskosiksi ja tämän päätössäännön perusteella heihin kohdistetaan erityistä huomioita terveydenhuollon toimenpiteiden osalta.

Havainnollistamme pikkukeskosmäärittelmään liittyvää epäjatkuvuutta sairaalahoidossa yhdistämällä Suomen syntymärekisteristä ja hoitoilmoitusrekisteristä saatuja tietoja sairaalapäivien määrästä ensimmäisen elinvuoden aikana suhteessa syntymäpainoon raskausviikoilla 32+ syntyneille (Kuvio 3A). Kuvassa sovitetaan lineaarinen käyrä keskimääräisten sairaalapäivien kokonaismäärille suhteessa syntymäpainoon epäjatkuvuuskohdan (1500g) molemmin puolin. Erot pikkukeskosten (alle 1500g) sovitteessa ja keskosten sovitteessa (yli 1500g) 1500 gramman rajalla kuvastavat sitä, että pikkukeskoshoido kasvatkaa vastasyntyneen sairaalassa vietettyjä päiviä keskimäärin 7 päivän verran. Tämä voi johtua siitä, että pikkukeskosten hoitoon tyyppillisesti liittyy tavanomaista todennäköisemmin mm. tehohoitoa (37).

Kyseenomainen epäjatkuvuus liittyy yksinomaan hallinnolliseen päätössääntöön eikä itse 1500g-rajaa liity lääketieteellisiä perusteita tai erityisiä eroja vanhempien taustatekijöissä kuten tulotasossa (Kuvio 3B).





Kuvio 3. Regressioepäjatkuuusmenetelmän hyödyntäminen pikkukeskoshoidon syntymäpainoon perustuvan raja-arvon puitteissa.

Seloste. Kuvan paneeleissa A ja B esitetään vastasyntyneen sairaalapäivät ja kotitalouden tulot regressioepäjatkuuusmenetelmän puitteissa tehdyllä kuviolla, jossa muodostetaan lineaarinen sovitte vasteelle (10 painokategoriaan perustuen) erikseen epäjatkuuuskohdan (1500g) molemmilta puolin painottaen kuitenkin rajan lähellä olevia havaintoja. Epäjatkuuuskohdan ero sovitteissa kuvastaa pikkukeskoshoidon vaikutusta. Tarkastelussa on hyödynnetty ns. “donitsi-RD” -menetelmää, ja poistettu tasan 1500 gramman havainnot. Paneelissa C esitetään syntymäpainojen frekvenssi 2.5 gramman välein.

Samanlaista pikkukeskohoitoon liittyvää päätössääntöä on toteutettu useassa kansainvälisessä tutkimuksessa. Syntymäpainoon liittyvää sairaalahoidon intensiteetin epäjatkuuutta on hyödynnetty pikkukeskoshoidojen vaikutusarvioinnissa Yhdysvaltojen (38,39), Norjan, Chilen (40) ja Tanskan (41) rekisteriaineistoihin perustuen. Näistä ensimmäinen, Douglas Almondin ja kumppanien tutkimus (38), kuuluu terveystaloustieteen moderneihin klassikoihin, ei ainoastaan artikkelin itsensä, vaan myös siihen liittyneen debatin (42-43) johdosta. Alkuperäisessä tutkimuksessa (38) arvioitiin regressioepäjatkuuusasetelmaan perustuen pikkukeskoshoidon alentavan imeväiskuolleisuutta 0.74 prosenttiyksikköä. Tätä seuranneessa tutkimuksessa (42) tuotiin kuitenkin esille huolia erityisesti painorajaan tuntumassa tehtävään pyörityksiin, joka johtaa havaintojen kasautumiseen tietyissä gramma- sekä unssilukemissa. Samantyyppistä ilmiötä on havaittavissa myös Suomessa. Kuvio 3C havainnollistaa, että pyörityksiä tapahtuu erityisesti sadan gramman mutta myös kymmenen ja viiden gramman tarkkuudella.

Kausaalipäätelyn osalta tämä on ongelmallista siinä tapauksessa, että pyörityminen ei ole satunnaista. Suosittu tapa ottaa huomioon ei-sattumanvaraiseen pyöritykseen liittyvät huolet on poistaa epäjatkuuusrajapinnan kohdalla olevat havainnot (esimerkissä 1500 gramman havainnot). Tämän ns. ”donitsi-RD:n” tuloksena alkuperäiset tulokset kuolleisuuteen

liittyvät estimaatit pienenevät keskimäärin 50 % (42), mutta pysyvät tilastollisesti merkitsevinä ja taloudellisesti merkityksellisinä. Vastineessaan Almond et al. (43) tekevät lisäanalyyskejä, ja esittävät, että 1500 gramman raja-arvoon liittyvä epäjatkuuus vastasyntyneiden kuolleisuudessa esiintyy erityisesti alhaisen laatuluokituksen sairaaloissa (mutta ei korkean laatuluokituksen sairaaloissa) myös siinä tapauksessa, että havainnot poistetaan 1500 gramman tuntumasta.

Tämä debatti havainnollistaa sitä, että itse hallinnollisen päätössääntöjen raja-arvon tuntumassa voi tapahtua pyöritystä tai suoranaista lukemien manipulaatiota, jotka tuottavat kausaalitulkintojen osalta huolia. Päätössääntöjen toteutuminen käytännössä ja niihin liittyvät lisätoimenpiteet saattavat myös vaihdella alueiden välillä. Niinpä regressioepäjatkuuusmenetelmän hyödyntämisen yhteydessä tuleekin ottaa huomioon useita kontekstispesifejä asioita kuten päätössääntöjen sitovuus, päätössääntöön liittyvän lisätoimenpiteiden vaihtelevuus sekä muut institutionaaliset seikat (esim. pikkukeskosten kohdalla surfaktanttihoito), jotka saattavat vaikuttaa erilaisiin lopputuloksiin eri konteksteissa. Pikkukeskoshoidon RD-tutkimusten uskottavuutta ovat vahvistaneet pohjoismaisin rekisteriaineistoin tehdyt jatkotutkimukset, joiden puitteissa on havaittu pikkukeskosdiagnosoin olevan yhteydessä korkeampiin peruskoulun päättödistuksen arvosanoihin samaisella 1500 gramman rajalla niin Chilessä kuin Norjassakin (28).

Edellä mainittujen huolenaiheiden lisäksi regressioepäjatkuuusmenetelmän tuloksia voi yleistää varsin rajoitetusti. Menetelmää hyödyntäen saadut tulokset pätevät ensisijaisesti aivan päätössäännön rajapinnassa, eikä välttämättä kaukana rajapinnasta. Edellä mainitusta esimerkiksi ei siis voida päätellä kuinka vaikuttavaa hoito olisi, jos pikkukeskosraja olisi esimerkiksi 2000g. Tässä mielessä kausaalipäätelmät toteutuvat regressioepäjatkuuuden kohdalla ainoastaan paikallisesti, tietyn kapean tarkasteluikkunan puitteissa (esim. yllä 1400g ja 1600g välillä). Koska tarkasteluikkuna on hyvin kapea, niin lisäesteenä suhteellisen pienten vaikutusten havaitsemisessa voi olla tilastollinen voima. Regressioepäjatkuuusmenetelmä kyseisessä esimerkissä toimiikin paremmin Suomea suuremmissa maissa, joissa vuodessa syntyy huomattavasti enemmän lapsia.

Suomalaisia terveystaloustieteeseen tai terveydenhuoltoon liittyviä tutkimuksia, jotka hyödyntävät regressioepäjatkuuusmenetelmää on toistaiseksi tehty tai ainakin julkaistu suhteellisen vähän. Yhdessä huomioarvoisessa suomalaisessa sovellutuksessa tarkastellaan julkisen suun terveydenhuollon saatavuuden laajentumisen vaikutusta hammashuollon käynteihin, hampaiden harjaamiseen ja hammasterveyteen (44). Tutkimuksessa hyödynnettiin sitä, että julkisin varoin tuettua hammashuoltoa laajennettiin vuodesta 1956 alkaen syntyneille, mutta ei sitä aiemmin syntyneille kohorteille. Tällöin vuosina 1955 ja 1956 syntyneet asettuivat eriarvoiseen asemaan suun terveydenhuollon saatavuuden osalta, kun muutoin ei ole syytä olettaa kohorttien eroavan toisistaan. Hyödyntäen regressioepäjatkuuusmenetelmää ja THL:n postikyselyaineistoa suun terveyden osalta, tutkijat tulkitsevat julkisen suun terveydenhuollon laajentamisen kasvattaneen hammashuollon käyttöä ja vähentäneen puuttuvien hampaiden määrää. Tämän lisäksi suomalaisessa (45) ja ruotsalaisissa (46) tutkimuksissa on viime aikoina hyödynnetty tarkkaa syntymäpäivämäärää tarkasteltaessa terveyskeskuslääkärikäyntien muutosta 18-vuotissyntymäpäivän kohdalla, jolloin nuorilta aletaan periä maksuja lääkärikäynneistä. Molemmista tutkimuksista hyödynnetään politiikkamuutoksia asiakasmaksujen perinnässä sekä arvioidaan näiden vaikutuksia hyödyntämällä regressio-epäjatkuuusmenetelmän ja erotusten erotus -menetelmän yhdistelmää (RD-DiD).

## INSTRUMENTTIMUUTTUJAMENETELMÄ

Kolmas taloustieteessä jo pidempään hyödynnetty kvasikokeellinen menetelmä on ns. instrumenttimuuttujamenetelmä (instrumental variable method, IV). Vaikka kyseinen menetelmä on teknisesti jonkin verran vaikeaselkoisempi, siinäkin perusideana on usein hyödyntää jotakin vahingossa tai sattuman kaltaisesti syntyvää vaihtelua mielenkiinnon kohteena olevassa politiikkatoimenpiteessä. Instrumenttimuuttujamenetelmä soveltuu erityisesti sellaisiin tilanteisiin, joissa ei ole käytettävissä tutkijan toimesta tai luonnostaan syntyneitä koe- ja verrokkiryhmiä.

Instrumenttimenetelmän lähtökohtana on oletus, että käytössä on lisämuuttuja, instrumentti, jonka avulla kiinnostuksen kohteena olevasta selittävästä muuttujasta (esim. asepalvelus) erotetaan sellainen vaihtelu, joka ei korreloi harhaa tuottavien mallista puuttuvien selittäjien kanssa. Lisäksi instrumenttimuuttujalla ei voi olla suoraa vaikutusta vasteeseen (esim. kuolleisuus), vaan se voi vaikuttaa pelkästään epäsuorasti mielenkiinnon kohteena olevan toimenpiteen kautta (ns. exclusion restriction). Jos tällainen muuttuja on saatavilla, voidaan kiinnostuksen kohteena olevan toimenpiteen vaikutus estimoida aineistosta luotettavasti.

Menetelmän käytön kynnyskysymys on siten löytää instrumenttimuuttuja, joka korreloi kiinnostuksen kohteena olevan toimenpiteen kanssa, mutta ei muiden havaitsemattomien tai mallista puuttuvien, vasteeseen vaikuttavien tekijöiden kanssa. Siten menetelmän luotettavuus nojaa pitkälti instrumenttimuuttujalta vaadittavien tiukkojen ehtojen tai oletusten uskottavuuteen. Instrumenttimuuttujamenetelmää voidaan teknisesti soveltaa myös ilman kokeelliseen tai kvasikokeelliseen vaihteluun perustuvaa instrumenttia, mutta käytännössä instrumentteja koskevia ehtoja on tällöin vaikeaa täyttää. Instrumenttimuuttujamenetelmän toinen rajoite on RD-menetelmän lailla sen rajoitettu yleistettävyyden tai lokaalisuus. Menetelmä voi tuottaa hieinan erilaisia tuloksia toimenpiteen vaikutukselle käytetystä instrumentista riippuen, koska menetelmässä hyödynnetään aina instrumentin aiheuttamaa variaatiota mielenkiinnon kohteena olevaan toimenpiteeseen.

Vaikka instrumenttimuuttujamenetelmää on hyödynnetty taloustieteessä jo lähes 100 vuoden ajan, aiemmin sitä käytettiin varsin mekaanisesti

kiinnittämättä tarpeeksi huomiota instrumenttimuuttujalta vaadittaviin vaativiin taustaoletuksiin. Taloustieteen Nobel-palkinnon saaneiden Joshua Angristin ja Guido Imbensin esittämät metodologiset laajennukset instrumenttimuuttujamenetelmään (47) sekä niihin liittyvät empiiriset sovellukset (48,49) ovat nostaneet menetelmän yhä keskeisempään rooliin kvasikokeellisia asetelmia hyödyntävässä empiirisessä tutkimuksessa.

Yhdessä ensimmäisistä moderneista IV-menetelmän sovelluksista tarkasteltiin Vietnamin sodan aikaisen asepalveluksen vaikutusta myöhemmän työuran ja ansioihin Yhdysvalloissa (48). Tutkimusasetelma perustui asepalveluksen todennäköisyyden satunnaiseen vaihteluun, joka syntyi syntymäajankohtaan perustuvan kutsuntajärjestyksen arvonnasta seurauksena. Kyseessä oli ensimmäinen sovellus, jossa ns. luonnollista koeasetelmaa, tässä tapauksessa kutsuntajärjestyksen arvontatulosta, hyödynnettiin instrumenttimuuttujana mielenkiinnon kohteena olevalle toimenpiteelle (asepalvelus). Myöhemmin samaa tutkimusasetelmaa on käytetty tutkimuksessa, jossa on arvioitu Vietnamin sodan aikaisen asepalveluksen vaikutusta kuolleisuuteen (49). Merkittävänä erona aiempiin IV-menetelmän sovelluksiin näissä tutkimuksissa oli nimenomaan se, että instrumentti perustui satunnaiseen arvontatuloksen luomaan vaihteluun.

Terveystaloustieteen sovelluksissa IV-menetelmää on hyödynnetty jo varsin pitkään (50), mutta vasta 2000-luvulla niin sanotun taloustieteen uskottavuusvallankumouksen myötä instrumentteina on alettu hyödyntämään erilaisia satunnaisia tai kvasisatunnaisia instrumenttimuuttujia interventioiden ja politiikkasäätöjen analyysissä (51,52). Esimerkiksi lääkärin, sosiaalityöntekijöiden tai muiden viranomaisten harkinnanvaraisiin päätöksiin perustuvat instrumentit ovat olleet suosittuja arvioitaessa erilaisten päätösten kuten opioidien määräämisen (53), tuetun asumisen (54), hoitokotiin pääsyn (55) tai lasten huostaanoton (56) vaikutuksia mielenkiinnon kohteena oleviin vasteisiin. Ideana näissä asetelmissa on hyödyntää yksilöiden kannalta lähes satunnaista vaihtelua viranomaisten tai muiden päätöksentekijöiden määräytymisessä palvelua käyttäville asiakkaille ja toisaalta viranomaisten välistä vaihtelua heille ominaisessa päätöksenteossa.

Suomalaisessa terveystaloustieteen tutkimuksessa instrumenttimuuttujamenetelmää on hyödynnetty muun muassa arvioitaessa keisarileikkauksen pidemmän aikavälin terveysvaikutuksia lapsiin (57). Keisarileikkauksen vaikutuksia tai syy-seuraussuhdetta lasten myöhempiin terveys-tulemiin on yleisesti hyvin haastavaa arvioida, koska keisarileikkauksella synnyttävät äidit ovat keskimäärin hyvin erilaisia alateitse synnyttäviin äiteihin verrattuna. Vaikka keisarileikkausten ja lasten terveystulemien yhteyksiä on tutkittu sadoissa tutkimuksissa maailmalla, löydettyjen yhteyksien avulla ei voida uskottavasti arvioida keisarileikkausten vaikutuksia. Tässä kyseenomaisessa suomalaisiin rekisteritietoihin perustuvasa tutkimuksessa hyödynnetään ajallista vaihtelua suunnittelemattomissa keisarileikkauksissa instrumenttimuuttujana havaituille keisarileikkauksille (57). Ideana on käyttää synnyttävien äitien ja lasten näkökulmasta lähes satunnaisesti syntyvää vaihtelua keisarileikkauksen todennäköisyydessä sen mukaan, minä viikonpäivänä lapsi on syntynyt suunnittelemattomalla sektiolla. Käytännössä tutkimuksessa hyödynnetään havaintoa siitä, että Suomessa suunnittelemattomia keisarileikkauksia tehdään virka-aikaan perjantaisin ja juuri ennen lomia todennäköisemmin kuin muina arkipäivinä. Tässä tapauksessa keisarileikkausten ajalliseen vaihteluun perustuva tutkimusasetelma mahdollistaa keisarileikkausten terveysvaikutusten arvioinnin varsin uskottavaan instrumenttimuuttujaan perustuen.

## **KVASIKOKEELLISTEN MENETELMIEN VAIKUTUKSET TERVEYSTALOUSTIETEESSÄ**

Miksi kvasikokeellisten asetelmien sekä niihin liittyvien ekonometristen menetelmien kehityksellä on sitten ollut niin merkittävä vaikutus empiiriseen tutkimukseen terveystaloustieteessä sekä kansainvälisesti että Suomessa? Taustalla on useita syitä, joista keskustelemme seuraavaksi.

Tärkein peruste liittyyne soveltavien tutkimusten uskottavuuden selkeään parantumiseen kvasikokeellisten asetelmien laajemman hyödyntämisen kautta. Vaikka monet kvasikokeelliset menetelmät (kuten regressioepäjatkuvuusmenetelmä) olivat jo olemassa tai niitä oli kehitetty muilla tieteenaloilla, niiden soveltaminen oli hyvin vähäistä 1990-luvun alkuun asti. Taloustieteen Nobelin vuonna 2021 voittaneiden David Cardin, Joshua Angristin ja Guido Imbensin

sekä monien muiden kvasikokeellisia asetelmia hyödyntäneet pioneiritutkimukset näyttivät esimerkkiä asetelmien eduista erilaisissa taloustieteellisissä sovelluksissa. Vaikka samoja kysymyksiä oli tutkittu aiemmin erilaisilla ekonometrisillä menetelmillä, tutkimusasetelmat alkoivat saada vähitellen enemmän painoarvoa empiirisissä sovelluksissa ekonometristen menetelmien sijasta. Tästä muutoksesta taloustieteen empiirisessä tutkimuksessa on käytetty nimeä uskottavuuden vallankumous (*credibility revolution*), jonka tienraivaajina pidetään juuri Cardia, Angristia ja Imbensia. Muutos alkoi työn taloustieteestä 1990-luvulla, mutta 2000-luvun aikana luonnollisia koeasetelmia vähitellen alettiin hyödyntämään yhä enemmän taloustieteen eri sovellusalueilla, mukaan lukien terveystaloustieteessä (1).

Kvasikokeellisten asetelmien yleistymisen toisena tärkeänä syynä voidaan pitää niiden yllättävänkin laaja soveltuvuus moniin tilanteisiin. Aiemmin ajateltiin, että kvasikokeellisia asetelmia voidaan hyödyntää vain hyvin rajatussa määrässä sovelluksia. Tähän ajatteluun oli syynä se, että asetelmien löytäminen vaatii usein tutkijoilta paljon instituutioihin ja järjestelmiin liittyvää selvitys- ja tutkimustyötä, eivätkä asetelmat ole useinkaan mitenkään ilmeisiä. Kvasikokeellisten asetelmien laajentunut hyödyntäminen eri taloustieteen osa-alueilla on kuitenkin paljastanut, että asetelmat soveltuvat moniin taloustieteen sovelluksiin erittäin hyvin. Myös terveystaloustieteessä kvasikokeellisten asetelmien laajempi hyödyntäminen näkyy selvästi esimerkiksi johtavien terveystaloustieteen lehtien (kuten *Journal of Health Economics*) viime vuosina julkaisemissa artikkeleissa (vrt. Taulukko 1). Erityisen huomionarvoista on ollut kvasikokeellisten asetelmien nopea hyödyntäminen koronapandemiaa koskevissa tutkimuksissa. Taloustieteilijät pystyivät ennakkojulkaisujen arkistojen kautta tuottamaan verratonta nopeasti päätöksenteon tueksi informaatiota mm. maskipakkojen (34) ja koulusulkujen (58) vaikutuksesta COVID-19 infektoiden leviämiseen sekä liikkumisrajoitusten mielenterveysvaikutuksista (59).

Yksi tärkeä käytännön syy kvasikokeellisten asetelmien suosioon on myös se, että asetelmien perusidea on yleensä hyvin yksinkertainen ja selitettävissä myös muille kuin tutkijoille. Kuten Angrist ja Pischke (51) artikkelissaan painottavat, tyypillisesti hyvät asetelmat mahdollistavat empiiristen menetelmien selittämisen yksinker-

taisella tavalla sekä tuloksien suoraviivaisen esittämisen. Esimerkiksi monissa kvasikokeellisia asetelmia hyödyntävissä sovelluksissa on vain kaksi ryhmää, toimenpiteen kohteena oleva koe-ryhmä sekä verrokkina toimiva kontrolliryhmä. Tällainen asetelma mahdollistaa yksinkertaiset ja intuitiiviset graafiset vertailut, joiden ymmärtäminen ei edellytä syvällistä osaamista tilastotieteestä tai ekonometriasta. Tästä johtuen asetelmia ja niihin perustuvia tutkimustuloksia voidaan esitellä ymmärrettävällä tavalla niin ensimmäisen vuoden yliopisto-opiskelijoille, medialle kuin virkamiehillekin. Asetelmien yksinkertaisuus on tehnyt tutkimuksista myös paljon läpinäkyvämpiä ja siten osaltaan mahdollistanut tutkimusten toistettavuutta ja uudelleenanalysointia kustannustehokkaasti. Lisäksi asetelmien keskeisten oletusten läpinäkyvyys on mahdollistanut oletusten uskottavuuden arvioinnin sekä erilaiset herkkyystarkastelut.

Kvasikokeellisten asetelmien hyödyntäminen on levinnyt vähitellen myös taloustieteen ulkopuolelle. Niitä käytetään nykyisin esimerkiksi valtio-opissa, sosiologiassa ja kasvatustieteessä. Vaikka epidemiologiassa ja yleisemmin lääketieteessä kvasikokeellisia asetelmia on hyödynnetty enenevässä määrin, kyseisillä tieteenaloilla on vielä runsaasti hyödyntämätöntä potentiaalia (60–62). Esimerkiksi instrumenttimuuttujamenetelmälle voisi olla paljon sovellusmahdollisuuksia erilaisissa lääketieteen satunnaistetuissa ja kvasikokeellisissa asetelmissa, joissa hoitotoimenpiteen noudattaminen on epätäydellistä.

## JOHTOPÄÄTÖKSET

Kvasikokeellisten menetelmien merkittävänä etuna voidaan pitää niiden soveltumista moniin tärkeisiin terveyspolitiikkaa ja terveydenhuoltoa käsitteleviin kysymyksiin ja lisäksi se, että menetelmät mahdollistavat uskottavien vaikuttavuus- ja kustannusvaikutusarviointien tekemisen vastaavasti kuin satunnaistetut koeasetelmat. Kvasikokeelliset menetelmät täydentävät terveystaloustieteen tutkimusta osa-alueilla, joihin kontrolloiduilla satunnaiskokeilla on haastavaa tai eettisesti mahdotonta vastata. Erityisesti niissä kysymyksissä, joihin ei ole saatavissa tietoa satunnaiskokeisiin perustuen, tulisi ensisijaisesti pyrkiä etsimään tietoa kvasikokeellisia menetelmiä hyödyntävistä tutkimuksista. Tästä syystä terveydenhuollon ammattilaisten koulutuksessa tulisi entistä laajemmin huomioida kvasikokeelliset me-

netelmät. Perusymmärrys kvasikokeellisten menetelmien taustaoletuksista, haasteista ja hyödyntämismahdollisuuksista auttaisi terveydenhuollon ammattilaisia luotettavan tutkimustiedon sekä hedelmällisten tutkimuskohteiden tunnistamisessa.

Terveydenhuollon tutkimuksen piirissä on paljon hyödyntämätöntä potentiaalia kvasikokeellisten asetelmien laajemmalle hyödyntämiselle. Kun otetaan huomioon, kuinka paljon satunnaistettujen koeasetelmien toteutus sitoo niukkoja voimavaroja, on jossain määrin yllättävää kuinka vähälle huomiolle varsin kustannustehokkaasti toteutettavat kvasikokeellisten menetelmien sovellutukset ovat tähän asti jääneet terveydenhuollon ammattilaisten piirissä. Taloustieteilijät ovat yhä enemmän siirtyneet tarkastelemaan erilaisiin terveyteen ja terveydenhuollon piiriin kuuluvia kysymyksiä kenties siitä syystä, että terveydenhuollon piiristä löytyy paljon hallinnollisia päätössääntöjä, jotka auttavat vastaamaan kysymyksiin, joihin aiempi tutkimus ei ole pystynyt tuottamaan tyydyttäviä vastauksia. Pelkän taloustieteen koulutuksen läpi käyneillä

taasen on paljon opittavaa terveyden instituutio-naalisista seikoista, erilaisista hallinnollisista päätössääntöistä ja niiden noudattamisesta sekä erilaisista epäjatkuvuuksista sosiaali- ja terveydenhuollon toiminnassa. Tältä osin näkisimme, että terveystaloustieteen ja lääketieteen koulutus- ja tutkimusyhteistyön piirissä on runsaasti hyödyntämätöntä potentiaalia.

#### KIITOKSET:

Haluamme kiittää Konsta Lavastetta sekä käsikirjoituksen arvioitsijoita kommenteista.

#### RAHOITTAJAT:

Tutkimusta ei ole rahoittanut mikään taho.

#### KIRJOITTAJIEN KONTRIBUUTIOT:

Kortelainen laati käsikirjoituksen suunnitelman. Kortelainen ja Salokangas osallistuivat käsikirjoituksen kirjoittamiseen ja kriittiseen kommentointiin. Salokangas keräsi taulukon aineistot ja muodosti kuvat. Molemmat kirjoittajat ovat hyväksyneet lopullisen käsikirjoituksen.

**Kortelainen, M., Salokangas, H. *Quasi-experimental methods in evaluating effectiveness in health care and health economics. Sosiaalilääketieteellinen aikakauslehti – Journal of Social Medicine* 2023; 60: 305–320.**

One of the most central concepts in health economics is effectiveness. The term ‘effect’ usually refers to how the outcome of interest in the treatment group under study develop after the intervention compared to the situation where the intervention would not have been carried out. In medicine, effectiveness is mostly studied using randomized trials. The role of randomized experiments has traditionally been less important in the social sciences, although large-scale social experiments have been carried out at least since the 1970s.

The added value of health economics as a tool for organizing healthcare operations has primarily been seen in the prioritization and systematic review of alternative uses of resources, not so much in the area of impact evaluation. However, the methodological advances in economics and their applications in health economics in recent decades have illustrated the important role of credible research designs in the evaluation of effectiveness. For many important questions in health economics, the most viable option for effectiveness evaluation is to use observational quasi-experimental designs and

related econometric methods instead of randomized experiments. A significant advantage of quasi-experimental methods is their applicability to many important issues related to health policy and healthcare services, and also, the fact that the methods enable credible effectiveness and cost-effectiveness evaluations to be made in a manner similar to randomized trial designs.

This article presents experimental and quasi-experimental methods used in modern health economics literature as well as some recent health economic applications based on quasi-experimental methods. In addition, the article evaluates the impact of the so-called credibility revolution on health economics research. Another aim of this paper is to illustrate for students and researchers the possibilities of utilizing quasi-experimental designs in topical questions dealing with social and health care.

**Keywords:** Causal inference, quasi-experimental design, natural experiment, health economics, effectiveness

Saapunut (25.10.2022)

Hyväksytty (02.08.2023)

- (1) Glied S. The Credibility Revolution in Economics and How It Has Changed Health Policy. *JAMA Health Forum* 2021;2:e214335. doi:10.1001/jamahealthforum.2021.4335
- (2) Kortelainen M. Taloustieteen Nobel luonnollisten koeasetelmien kehittäjille ja hyödyntäjille. *KAK* 2021;117;568–576. [https://www.taloustieteellinenyhdistys.fi/wp-content/uploads/2021/12/KAK\\_4\\_2021\\_WEB-92-100.pdf](https://www.taloustieteellinenyhdistys.fi/wp-content/uploads/2021/12/KAK_4_2021_WEB-92-100.pdf)
- (3) Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *J Am Stat Assoc* 2005;100:322–331. doi:10.1198/016214504000001880
- (4) Sacks H, Chalmers TC, Smith HJ. Randomized versus historical controls for clinical trials. *Am J Med* 1982;72:233–240. doi: 10.1016/0002-9343(82)90815-4
- (5) Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New Engl J Med* 2000;342:1887–1892. doi: 10.1056/NEJM200006223422507
- (6) Rothwell PM. External validity of randomised controlled trials: “to whom do the results of this trial apply? *Lancet* 2005;365:82–93. doi: 10.1016/S0140-6736(04)17670-8
- (7) Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 2018;210:2–21. doi: <https://doi.org/10.1016/j.socscimed.2017.12.005>
- (8) Jones DS, Podolsky SH. The history and fate of the gold standard. *Lancet* 2015;385:1502–1503. doi: [https://doi.org/10.1016/S0140-6736\(15\)60742-5](https://doi.org/10.1016/S0140-6736(15)60742-5)
- (9) Feinstein AR, Horwitz RI. Problems in the “evidence” of “evidence-based medicine”. *Am J Med*, 1997;103:529–535. doi: 10.1016/s0002-9343(97)00244-1
- (10) Newhouse JP. Free for all? Lessons from the RAND health insurance experiment. *HUP*; 1993. doi: <https://doi.org/10.7249/CB199>
- (11) Finkelstein A, Taubman S, Wright B ym. The Oregon health insurance experiment: evidence from the first year. *Q J Econ* 2012;127:1057–1106. doi: <https://doi.org/10.1093/qje/qjs020>
- (12) Sääksvuori L, Betsch C, Nohynek H, ym. Information nudges for influenza vaccination: Evidence from a large-scale cluster-randomized controlled trial in Finland. *Plos Med* 2022; in press. doi: <https://doi.org/10.1371/journal.pmed.1003919>
- (13) Cunningham S. Causal inference: The mixtape (kappale 9.6.9). Yale university press, 2021.
- (14) Snow J. On the mode of communication of cholera. 2. painos. London: John Churchill; 1855.
- (15) Thistlewaite DL, Campbell DT. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *J Educ Psychol* 1960;51:309–317. doi: <https://doi.org/10.1037/h0044319>
- (16) Campbell DT. Reforms as experiments. *Am Psychol* 1969;24:409–429. doi: <https://doi.org/10.1037/h0027982>
- (17) Huntington-Klein, N. The effect: An introduction to research design and causality. Chapman and Hall/CRC;2021. <https://theeffectbook.net/>
- (18) Vaalavuo M, Salokangas H, Tahvonen O. Gender inequality reinforced: the impact of a child’s health shock on parents’ labor market trajectories. *Demography*. 2023;4. doi: <https://doi.org/10.1215/00703370-10828906>
- (19) Jeon SH, Pohl RV. Health and work in the family: Evidence from spouses’ cancer diagnoses. *J Health Econ* 2017;52:1–18. doi: <https://doi.org/10.1016/j.jhealeco.2016.12.008>
- (20) Dobkin C, Finkelstein A, Kluender R, ym. The economic consequences of hospital admissions. *Am Econ Rev* 2018; 108: 308–352. doi: 10.1257/aer.20161038
- (21) Fadlon I, Nielsen TH. Family health behaviors *Am Econ Rev* 2019;109:3162–3191. doi: 10.1257/aer.20171993
- (22) Ahomäki I, Pitkänen V, Soppi A, ym. Impact of a physician-targeted letter on opioid prescribing. *J Health Econ* 2020;72: 102344. doi: <https://doi.org/10.1016/j.jhealeco.2020.102344>
- (23) Böckerman P, Kortelainen M, Laine L, ym. Digital waste? Unintended consequences of health information technology. *VATT WP* 2019;117. <http://dx.doi.org/10.2139/ssrn.3390265>
- (24) Böckerman P, Laine L, Nurminen M, ym. Information integration, coordination failures, quality of prescribing. *J Hum Resour julkaisuprosessissa*. doi: 10.3368/jhr.0921-11910R2
- (25) Kortelainen M, Lavaste K, Laine L, ym. Hospital choice, allocation, and performance: Evidence from a reform in the public health care system. 2023.
- (26) Kortelainen M, Markkanen J, Siikanen M, ym. The effects of price regulation on pharmaceutical expenditure and availability. 2023. Helsinki GSE Discussion Papers 16/2023. <https://www.helsinki.gse.fi/discussion-papers/discussion-paper-16>
- (27) Haaga T, Böckerman P, Kortelainen M, ym. Effects of nurse visit copayment on primary care use: Do low-income households pay the price? 2023. <https://osf.io/skuv9/>
- (28) Goodman-Bacon A. Difference-in-differences with variation in treatment timing. *J Econom* 225.2 (2021): 254–277. doi: 10.1016/j.jeconom.2021.03.014

- (29) Callaway B, Sant'Anna, P. Difference-in-differences with multiple time periods. *J Econom* 2021;225: 200–230. doi: 10.1016/j.jeconom.2020.12.001
- (30) Roth J, Sant'Anna P. Efficient estimation for staggered rollout designs. *J Pol Econ: Microecon, julkaisuprosessissa* 2023. doi: <https://doi.org/10.1086/726581>
- (31) Cengiz D, Dube A, Lindner A, ym. The effect of minimum wages on low-wage jobs. *Q J Econ* 2019;134: 1405–1454. doi: <https://doi.org/10.1093/qje/qjz014>
- (32) Abadie A. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *J Econ Lit* 2021;59:391–425. doi:10.1257/jel.20191450
- (33) Arkhangelsky D, Athey S, Hirshberg DA, ym. Synthetic difference in differences. *Am Econ Rev* 2021;111:4088–4118. doi: 10.1257/aer.20190159
- (34) Mitze T, Kosfeld R, Rode J, ym. Face masks considerably reduce COVID-19 cases in Germany. *PNAS*, 2020;117:32293–32301. doi: <https://doi.org/10.1073/pnas.2015954117>
- (35) Hämäläinen K, Moisio A. One or two tiers of local government? The cost effects of regional experiment. *VATT WP* 2015:65.
- (36) Heckley G, Nordin M, Gerdtham UG. The health returns of attending university for the marginally eligible student. *Health Econ* 2022;31:877–903 doi: <https://doi.org/10.1002/hec.4484>
- (37) Lehtonen L. Keskosen muuttuva hoito. *Duodecim* 2009;125:1333–9. <https://www.duodecimlehti.fi/duo98131>
- (38) Almond D, Doyle JJ, Kowalski A, ym. Estimating marginal returns to medical care: Evidence from at-risk newborns. *Q J Econ* 2010;125:591–634. doi: <https://doi.org/10.1162/qjec.2010.125.2.591>
- (39) Chyn E, Gold S, Hastings J. The returns to early-life interventions for very low birth weight children. *J Health Econ*. 2021;75:102400. doi: <https://doi.org/10.1016/j.jhealeco.2020.102400>
- (40) Bharadwaj P, Løken KV, Neilson C. Early life health interventions and academic achievement. *Am Econ Rev* 2013;103:1862–91. doi: 10.1257/aer.103.5.1862
- (41) Daysal NM, Simonsen M, Trandafir M, ym. Spillover effects of early-life medical interventions. *Rev Econ Stat* 2022;104:1–16. doi: [https://doi.org/10.1162/rest\\_a\\_00982](https://doi.org/10.1162/rest_a_00982)
- (42) Barreca AI, Guldi M, Lindo JM, ym. Saving babies? Revisiting the effect of very low birth weight classification. *Q J Econ* 2011;126:2117–2123. doi: <https://doi.org/10.1093/qje/qjr042>
- (43) Almond D, Doyle JJ, Kowalski A, ym. The role of hospital heterogeneity in measuring marginal returns to medical care: a reply to Barreca, Guldi, Lindo, and Waddell. *Q J Econ* 2011;126:2125–2131. doi: <https://doi.org/10.1093/qje/qjr037>
- (44) Raittio E, Suominen AL. Effects of universal oral healthcare coverage in an adult population: A long-term nationwide natural experiment. *Community Dent Oral* 2022 (in press). doi: 10.1111/cdoe.12785
- (45) Haaga T, Böckerman P, Kortelainen M, ym. Do low-income households respond more to cost sharing in primary care? Evidence from regression discontinuity design, 2023. <https://osf.io/4c8af>
- (46) Nilsson A, Paul A. Patient cost-sharing, socioeconomic status, and children's health care utilization. *J Health Econ*. 2018;59:109–124. doi: <https://doi.org/10.1016/j.jhealeco.2018.03.006>
- (47) Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica* 1994;61:467–476. doi: <https://doi.org/10.2307/2951620>
- (48) Angrist JD. Lifetime earnings and the Vietnam era draft lottery: Evidence from social security administrative records. *Am Econ Rev* 1990;80:313–385. <https://www.jstor.org/stable/2006669>
- (49) Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996;91:444–455. doi: <https://doi.org/10.2307/2291629>
- (50) Keele L, Small D. Instrumental variables: Don't throw the baby out with the bathwater. *Health Serv Res*. 2019;54:543–546. doi: 10.1111/1475-6773.13130
- (51) Angrist JD, Pischke JS. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *J Econ Perspect* 2010;24:3–30. doi: 10.1257/jep.24.2.3
- (52) Cawley J. A selective review of the first 20 years of instrumental variables models in health-services research and medicine. *Journal of medical economics*. 2015;18:721–734. doi: <https://doi.org/10.3111/13696998.2015.1043917>
- (53) Eichmeyer S, Zhang J. Pathways into Opioid Dependence: Evidence from Practice Variation in Emergency Departments. *Am Econ J: Appl Econ*, 2022;14:271–300. doi: 10.1257/app.20210048
- (54) Lopes FV, Bakx P, Harper S, ym. The effects of supported housing for individuals with mental disorders. *Health Econ* 2022;31:115–133, doi: <https://doi.org/10.1002/hec.4579>
- (55) Bakx P, Wouterse B, van Doorslaer E, ym. Better off at home? Effects of nursing home eligibility on costs, hospitalizations and survival. *J Health Econ*, 2020;73:102354, doi: <https://doi.org/10.1016/j.jhealeco.2020.102354>
- (56) Doyle JJ. Child protection and child outcomes: Measuring the effects of foster care. *Am Econ Rev* 2007;97:1583–1610. doi: 10.1257/aer.97.5.1583
- (57) Costa-Ramón A, Kortelainen M, Rodríguez-González A, ym. The long-run effects of cesarean

- sections. *J Hum Resour* 2022;0719-10334R1.  
doi: 10.3368/jhr.58.2.0719-10334R1
- (58) Vlachos J, Hertegård E, Svaleryd H. The effects of school closures on SARS-CoV-2 among parents and teachers. *PNAS* 2021;118:e2020834118.  
doi: <https://doi.org/10.1073/pnas.2020834118>
- (59) Altındag O, Erten B, Keskin P. Mental health costs of lockdowns: Evidence from age-specific curfews in Turkey. *Am Econ J-appl Econ* 2022;14:320–43.  
doi: 10.1257/app.20200811
- (60) Moscoe E, Bor J, Bärnighausen JT. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: A review of current and best practice. *J Clin Epidemiol* 2015;68:122–133.  
doi: 10.1016/j.jclinepi.2014.06.021
- (61) Craig P, Katikireddi SV, Leyland A, ym. Natural experiments: An overview of methods, approaches, and contributions to public health intervention research. *Annu Rev Publ Health* 2017;38: 39–56.  
doi: 10.1146/annurev-publhealth-031816-044327
- (62) Hilton Boon M, Craig P, Thomson H, ym. Regression discontinuity designs in health: A systematic review. *Epidemiology* 2021;32:87–93.  
doi: 10.1097/EDE.0000000000001274

MIKA KORTELAINE  
*KTU, Professori*  
*Turun yliopisto ja THL*

HENRI SALOKANGAS  
*VTT, Erikoistutkija*  
*THL*