

Sosiaali- ja terveystalvelujen vaikuttavuuden ja kustannusvaikuttavuuden tutkiminen

Janissa Miettinen, Kirsikka Selander & Ismo Linnosmaa

Johdanto

Terveystalveluollossa talvelujen ja menetelmien vaikuttavuuden ja tuloksellisuuden arvioinnilla on pitkät perinteet, kun taas sosiaali-talvelujen vaikuttavuuden arviointiin liittyvä keskustelu on selvästi uudempi ilmiö (Rajavaara 2007; Kivipelto, Blomgren, Karjalainen & Saikkonen 2013). Vaatimukset esimerkiksi sosiaalityön vaikuttavuuden mittaamiseksi ovat lisääntyneet (Sosiaalihuollon lainsäädännön uudistamistyöryhmä 2012). Tämä johtuu muun muassa siitä, että vaikuttavuuden arvioinnista syntyvää tietoa pidetään edellytyksenä talvelujärjestelmän kehittämislle. Taustalla vaikuttaa myös lainsääd-

däntö sekä siinä tapahtuneet muutokset. Terveydenhuoltolain mukaan terveydenhuollon toiminnan on oltava laadukasta, turvallista ja näyttöön perustuvaa (L1326/2010, 8 §). Vastaavasti Sosiaalihuoltolain (L1301/2014, 15 §) mukaan sosiaalityön vaikuttavuutta tulee seurata.

Sosiaalialalla näyttöön perustuva työskentelyote jakaa kuitenkin vielä mielipiteitä. Lisäksi sosiaalihuollon toimintaympäristö on usein terveydenhuollon toimintaympäristöä monimutkaisempi, mikä heijastuu palvelujen laadun ja vaikuttavuuden seurantaan ja arviointiin (ks. Ikonen & asiantuntijaryhmä 2019, 11). Vaikuttavuustiedon soveltamisen vakiinnuttaminen eli näyttöön perustuvan ajattelun ja toiminnan edistäminen sosiaali- ja terveydenhuollon palvelujärjestelmän integroitua edellyttää sosiaali- ja terveystieteiden johtajilta ammattirajat ylittävää keskusteluyhteyttä ja monialaista käytäntöjen kehittämisjohtamista (Lammintakanen, Rissanen, Peronmaa-Hanska, Joensuu & Ruottu 2016, 5, 33). Lisäksi tarvitaan menetelmällistä perusymmärrystä sosiaali- ja terveystieteiden palvelujen vaikuttavuus- ja kustannusvaikuttavuustutkimuksen periaatteista ja erityyppisen vaikuttavuustiedon sovellettavuudesta käytännössä.

Laatu- ja vaikuttavuustiedon merkitys tiedolla johtamisen välineinä kasvaa (Ikonen & asiantuntijaryhmä 2019, 10). Palvelujen vaikuttavuuden arviointia ja vaikuttavuustiedon hyödyntämistä voidaan perustella sosiaali- ja terveydenhuollon käytännöistä kumpuavilla syillä. Terveydenhuollon tavoitteina on muun muassa edistää ja ylläpitää väestön ja yksilöiden terveyttä, työ- ja toimintakykyä, sosiaalista turvallisuutta sekä kaventaa terveyseroja (Terveydenhuoltolaki L1326/2010). Vastaavasti sosiaalihuollon tavoitteina on edistää ja ylläpitää hyvinvointia, sosiaalista turvallisuutta ja osallisuutta sekä vähentää eriarvoisuutta (Sosiaalihuoltolaki L1301/2014). Luotettava tieto toiminnan vaikuttavuudesta auttaa sosiaali- ja terveydenhuollon organisaatioiden johtajia ja henkilöstöä toteuttamaan näitä organisaatioiden toiminnalle asetettuja perustavoitteita. Tieto laadukkaista ja vaikuttavista palveluista on tärkeää myös siksi, että se auttaa sosiaali- ja terveydenhuollon päätöksentekijöitä kohdentamaan rajal-

liset resurssit mahdollisimman tehokkaasti niihin palveluihin ja toimenpiteisiin, joista on saatavilla mahdollisimman paljon terveyttä ja hyvinvointia. Tämän edistämiseksi viime vuosina onkin keskusteltu kansallisen laaturekisterin perustamisesta. Sen avulla sosiaali- ja terveyspalvelujen laatua voitaisiin seurata järjestelmällisesti ja luotettavasti, sekä vertailla laatua eri järjestäjien välillä (Jonsson, Pikkujämsä & Heliö 2019). Näin sosiaali- ja terveysalan toimijat voisivat päästä myös palvelujen vaikuttavuuden jäljille, vaikkei laatu itsessään olekaan tae palvelun vaikuttavuudesta.

Tässä luvussa keskitymme tarkastelemaan asioita, joita on huomioitava tutkittaessa sosiaali- ja terveyspalvelujen vaikuttavuutta ja joista vaikuttavuustiedon soveltajien on hyvä olla tietoisia. **Palveluilla** tarkoitamme eriasteista puuttumista (interventio) potilaiden tai asiakkaiden elämäntilanteisiin. Palvelut voivat olla tarkkaan määriteltävissä olevia toimenpiteitä ja/tai sosiaali- ja terveydenhuollon ammattilaisen harkintaan, taitoihin ja tietoperustaan perustuvia joustavasti palveluprosessin aikana määriteltäviä ja yhdessä potilaan tai asiakkaan kanssa toteutettavia muutosstrategioita. (Fraser 2009, 5, 18.)

Määrittelemme aluksi vaikuttavuuden käsitteen. Sen jälkeen tarkastelemme erilaisia mittareita, joita on mahdollista soveltaa vaikuttavuuden tutkimiseen. Kun vaikuttavuuden mittaamisen kysymykset on ratkaistu, on valittava tutkimuskysymykseen soveltuva **tutkimusasetelma** eli määriteltävä tutkimusaineisto, sen keruu ja tutkimuksessa käytettävät tutkimus- ja analyysimenetelmät. Palvelujen vaikuttavuuden todentamiseen on käytössä sekä kokeellisen että havainnoivan aineiston keruuseen perustuvia tutkimusasetelmia. Esittelemme nämä tutkimusasetelmat ja arvioimme niiden vahvuuksia ja heikkouksia luotettavan vaikuttavuustiedon tuottamisessa. Lopuksi kerromme, kuinka palvelujen kustannusvaikuttavuutta voidaan arvioida ja pohdimme vaikuttavuus- ja kustannusvaikuttavuustiedon käyttöä uudistuvan sosiaali- ja terveysalan kehittämisessä.

Vaikuttavuustiedon tuottamista ja hyödyntämistä ovat vähentäneet useat seikat. Ensinnäkin luotettavan tutkimustiedon tuottaminen palvelujen ja toimenpiteiden vaikuttavuudesta on hidasta. Sosiaali- ja terveydenhuollon johtajien ja päätöksentekijöiden näkökulmasta tutkimustulokset valmistuvat siten yleensä liian myöhään, ajatellen esimerkiksi tiettyjen asiakas- tai potilasryhmien palvelujen laatu- ja vaikuttavuustiedon vertailua käytännön päätöksentekotilanteissa. Vaikuttavuustiedon käyttöä on vähentänyt myös se, että tieto palvelujen vaikuttavuudesta ei ole kootusti ja nopeasti saatavilla. Lisäksi vaikuttavuustiedon tuottamiseen, analysointiin ja tulkinnaan ei ole ollut saatavilla suomenkielisiä lähteitä. Tämän artikkelin avulla pyrimme vastaamaan erityisesti tähän viimeiseen tarpeeseen.

Tämän artikkelin tavoitteena on antaa kokonaiskuva sosiaali- ja terveyspalvelujen vaikuttavuuden tutkimisen perustiedosta ja tutkimuksen erilaisista vaihtoehdoista. Artikkelin painopiste on menetelmällinen, sillä tarkoitus on auttaa vaikuttavuustutkimusta suunnittelevia tutkijoita ja opiskelijoita vaikuttavuus- ja kustannusvaikuttavuustutkimukseen kohdistuvien ratkaisujen tekemisessä. Samalla artikkeli tarjoaa sosiaali- ja terveysalan johtajille ja päätöksentekijöille tietopaketin vaikuttavuustiedon muodostumisesta, minkä tarkoituksena on auttaa heitä hoito- ja palveluvaihtoehtoja koskevan tutkimusnäytön arvioinnissa ja vertailussa.

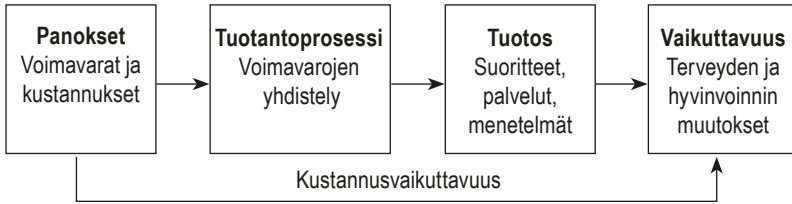
Vaikuttavuuden käsite ja lähestymistapa vaikuttavuuden arviointiin

Vaikuttavuudella tarkoitamme Sintosen ja Pekurisen (2006, 53) määritelmän mukaan toiminnan vaikutuksesta tapahtuvaa muutosta toiminnan perimmäisessä tavoitteessa. Vaikuttavuuden käsite määritellään siis suhteessa toiminnalle määriteltyihin tavoitteisiin, jotka voidaan edelleen liittää sosiaali- ja terveyspalvelujen perimmäisiin tavoitteisiin. Esimerkiksi Paasio (2006, 99) arvottaa julkisen sektorin

tärkeimmäksi tehtäväksi ihmisten hyvinvoinnin edistämisen. Sosiaali- ja terveydenhuollon asiayhteydessä toiminnan vaikuttavuudella tarkoitetaan siis palvelun tai muun toimenpiteen aiheuttamaa muutosta asiakkaan hyvinvoinnissa tai potilaan terveydentilassa.

Vaikuttavuudella on monia lähikäsitteitä, kuten tuloksellisuus, laatu ja tehokkuus, mutta vaikuttavuuden käsitteen ymmärtämiseksi on erityisen tärkeää ymmärtää, kuinka se eroaa vaikutuksen käsitteestä. Esimerkiksi Vedung (1997) on jäsentänyt vaikutuksia hyvin erilaisiin mahdollisiin tapauksiin: palvelun vaikutukset voivat olla esimerkiksi tavoiteltuja, toivottuja ja odotettuja tai tavoittelemattomia ja ei-toivottuja tai odottamattomia palvelun käytöstä syntyviä seurauksia. Kaksi asiaa voi siis olla kausaalissa suhteessa keskenään, vaikka kyse ei olekaan vaikuttavuudesta; tällöin on kyse yleisemmin asioiden välisistä syy-seuraussuhteista (Paasio 2006, 102). Sosiaali- ja terveyspalveluilla voi olla erilaisia toiminnasta johtuvia vaikutuksia, jotka eivät välttämättä kuitenkaan liity toiminnalle määriteltäviin tavoitteisiin tai palvelujen perimmäiseen tarkoitukseen. Vaikuttavuuden käsite eroaa vaikutuksista siten, että vaikuttavuus määritellään suhteessa tarkasteltavan toiminnan tavoitteista johdettuihin vaikuttavuuskriteereihin.

Vaikuttavuuden käsitettä voidaan edelleen syventää panos-tuotos-lopputulostuotosten mallin kautta (kuvio 1). Tämä terveystaloustieteellisistä lähtökohdista muotoiltu mallinnus lähtee liikkeelle panoksista, joilla tarkoitetaan sosiaali- ja terveyspalvelujen tuottamiseen tarvittavia voimavaroja, ja jatkuu tuotantoprosessiin, jossa panoksia yhdistellään tuotosten eli palvelujen, suoritteiden ja menetelmien aikaansaamiseksi. Sosiaali- ja terveyspalvelujen voidaan sanoa olevan vaikuttavia, kun palvelut edistävät asiakkaiden terveyttä ja/tai hyvinvointia eli tuottavat palveluille asetettujen tavoitteiden mukaisia vaikutuksia. Kun tunnetaan palvelujen kustannukset ja vaikuttavuus, voidaan sopivalla tutkimusasetelmalla tutkia myös palvelujen **kustannusvaikuttavuutta**.



Kuvio 1. Vaikuttavuuden ja kustannusvaikuttavuuden käsite
(mukaillen Sintonen & Pekurinen 2006, 52)

Kuvio 1 kuvaa prosessin, jonka kautta tuotettujen palvelujen vaikuttavuus muodostuu. Tätä prosessia ja erityisesti palvelun vaikuttavuuden syntymekanismeja tutkitaan niin sanotussa teoriaperustaisessa (*theory-based*) vaikuttavuuden arvioinnin lähestymistavassa (esim. Dahler-Larsen 2005) ja sen alasuuntauksissa, kuten realistisen arvioinnin piirissä (esim. Pawson & Tilley 1997)¹. Tässä artikkelisssa vaikuttavuuden arviointia lähestytään kuitenkin soveltaen tavoite- ja tarveperusteisen arvioinnin näkökulmia (ks. Rossi, Freeman & Lipsey 2004, luvut 7–9: *impact assessment* -lähestymistapa). Toisin sanoen tarkastelemalla, täyttyivätkö määritellyt palvelutarpeet tai tapahtuivatko odotetut terveyden ja hyvinvoinnin muutokset.

Vaikuttavuuskäsitteen määritelmään liittyy myös tutkimusasetelma, josta käsin palvelujen vaikuttavuutta tutkitaan. Vaikuttavuudesta (*effectiveness*) puhutaan yleensä silloin, kun tutkitaan sosiaali- ja terveydenhuollon tavanomaisen toiminnan terveys- ja hyvinvointivaikutuksia. Vuorela, Malmivaara, Komulainen ja Jousilahti (2014) käyttävät tässä yhteydessä myös ”arkivaikuttavuuden” käsitettä. **Te-hon** käsitettä (*efficacy*) taas käytetään viitattaessa ihanteellisissa olosuhteissa, kuten laboratoriossa, suoritettujen, tarkkaan valvottujen kliinisten kokeiden (satunnaistetut ja kontrolloidut tutkimusasetel-

¹ Palvelujen vaikuttavuuden tutkimisen ohella niin sanotun ohjelmateorian kehittäminen ja arvioiminen on myös tärkeää sen tutkimiseksi, miten ja millä edellytyksillä palvelulla voidaan saavuttaa tavoitteiden mukaisia vaikutuksia (ks. lisätietoja esim. Rossi ym. 2004; Chen 2005; Dalher-Larsen 2005).

mat) tutkimustuloksiin.² Tässä artikkelissa emme erottele vaikuttavuuden ja tehon käsitteitä vaan puhumme yleisesti vaikuttavuudesta, joka on vaikuttavuustutkimukseen vakiintunut suomenkielinen peruskäsite.

Vaikuttavuuden mittaaminen

Vaikuttavuustutkimuksen suunnittelu alkaa tutkimuskohteen määrittelystä sekä tutkimuskysymysten ja tutkimushypoteesien muotoilusta (Virtanen 2007; Kirk 2013, 2). Tällöin määritellään muun muassa se, mitataanko vaikuttavuutta yleisillä järjestelmä- tai väestötasoisilla mittareilla vai asiakasepisoditasoisilla, ongelmaspesifeillä mittareilla (Torkki, Leskelä, Linna, Torvinen, Klemola, Sinivuori, Larsio & Hörhammer 2017, 38). Kun arvioinnin kohde on tiedossa ja tiedetään, mitä tavoiteltavat vaikutukset ovat, tehdään päätökset ja määritelmät tutkimuksen tulosmuuttujista ja mitä **tulosmittareita** (*outcome measure(s)*) niiden mittaamiseen käytetään. Nämä tulosmittarit voidaan jakaa yksilöön liittyviin mittareihin kuten PROM (*Patient Reported Outcome Measures, PROM*) ja PREM (*Patient Reported Experience Measures, PREM*) -mittareihin, sekä erilaisiin rekisteritietoja tai tietokantoja hyödyntäviin objektiivisiin mittareihin. Näistä rekistereihin perustuvat mittarit ovat olleet Suomessa selkeästi käytetympiä, mutta viime aikoina myös erilaisten potilaan ja asiakkaan terveyttä ja toimintakykyä tarkastelevien mittareiden käyttö on yleistynyt. Tässä artikkelissa keskitymme erityisesti itseraportoituihin mittareihin, mutta sivuamme lopuksi myös erilaisiin tietojärjestelmiin ja rekistereihin kerättävään tietoon pohjautuvia tulosmittareita (*performance outcome, PerfO*). Lopuksi kerromme lyhyesti,

² Tehon ja vaikuttavuuden käsitteet tulevat ymmärrettäviksi suhteessa toisiinsa esimerkiksi ehkäisevien palvelujen tutkimusyyklin (*preventive intervention research cycle*) kautta (ks. Gottfredson, Cook, Gardner, Gorman-Smith, Howe, Sandler & Zafft 2015); artikkeli toimii hyvänä lisätiedon lähteenä tehon ja vaikuttavuuden käsitteisiin.

miten yksilötasolta voitaisiin edetä tarkastelemaan esimerkiksi koko palvelujärjestelmän vaikuttavuutta.

Itseraportoidut tulostittarit

Palvelun vaikuttavuutta voidaan mitata tiedustelemalla potilaan tai asiakkaan omaa arviota omasta toimintakyvystään tai elämän laadusta. Tällöin terveydenhuollon kontekstissa puhutaan potilaan raportoimasta vaikuttavuudesta (*Patient-Reported Outcomes, PRO*), kun taas sosiaalipuolella viitataan asiakkaan (*Consumer-reported Outcome measures, CROM*) kokemukseen. Vastaavasti jos halutaan tietää potilaan tai asiakkaan tyytyväisyydestä ja palvelukokemuksista, käytetään erilaisia PREM-mittareita (*Patient Reported Experience Measures*). Aina potilas tai asiakas ei ole kuitenkaan kykenevä arvioimaan omaa tilannettaan johtuen esimerkiksi vanhuudesta, päih-teistä tai vammasta. Sen vuoksi itseraportoituihin tulostittareihin voidaan sisällyttää myös erilaiset ammattilaisten (*clinician-reported outcomes, ClinRO*) tai asiakkaan läheisen (*observer reported outcome, ObsRO*) tekemät arviot potilaan tilanteesta. (Walton, Powers, Hobart, Patrick, Marquis, Vamvakas, Isaac, Molsen, Cano & Burke 2015; Torkki ym. 2017, 37–42.) Valintaa eri tason mittareiden välillä ohjaavatkin tutkimuksen kohderyhmä, tutkittava palvelu ja sen odotettavissa olevien vaikutusten luonne (vaikutukset esim. terveyteen, toimintakykyyn tai kustannuksiin).

Asiakkaan itsensä raportoimat mittarit (PROM ja CROM) ovat niin sanottuja sateenvarjokäsitteitä, jotka viittaavat sekä yhdestä muuttujasta että useista ulottuvuuksista koostuviin itsearviointimittareihin. Tällaisia ovat esimerkiksi potilaan itsensä tekemän arviot oireistaan, toimintakyvystään tai elämänlaadusta (*quality of life, QOL*).

Elämänlaadun arviota voidaan hyödyntää terveydenhuollon kontekstissa esimerkiksi silloin, kun halutaan tietää, miten vähäiset terveydenhuollon resurssit saataisiin kohdennettua siten, että niillä saa-

vutettaisiin mahdollisimman paljon terveyttä (esim. Williams 1996). Tällöin arvioidaan niin sanottujen **laatupainotettujen elinvuosien** (QALY, *Quality Adjusted Life Years*) määrää eli terveydenhuollon menetelmien tai palvelujen aiheuttamaa lisäystä elinvuosien määrässä (rekisteritietoihin/tutkimukseen perustuva arvio) ja elämän laadussa (itseraportoituna) (esim. Williams 1996; Weinstein, Torrance & McGuire 2009; Whitehead & Ali 2010; kritiikistä esim. Prieto & Sacristán 2003, ks. Esimerkki 1).

Esimerkki 1. Mitä QALY mittaa?

QALY mittaa terveydenhuollon palveluiden vaikutuksia potilaan elinikään ja elämän laatuun. Elinikää mitataan elinvuosina ja elämänlaatua mittareilla, joiden arvot vaihtelevat yleensä välillä 0–1. Arvo 1 on paras mahdollinen terveydentila ("täysin terve"), kun taas arvo 0 viittaa potilaan kuolemaan. Jos potilas saa hoitojen seurauksena esimerkiksi yhden elinvuoden lisää täydellä elämän laadulla, hoito tuottaa potilaalle yhden QALY:n. Sen sijaan, jos hoito tuottaa yhden lisäelinvuoden ja potilaan elämänlaatu on 0,5 tänä aikana, on muutos puolikkaan QALY:n arvoinen. Vastaavasti, jos hoito lisää potilaan elinikää puolella vuodella ja potilas elää tämän ajan täysin terveenä, on kyse jälleen puolikkaan QALY:n arvoisesta muutoksesta. QALY:jen muutokset voivat olla myös negatiivisia, mikäli toimenpiteet vähentävät laatupainotettuja elinvuosia.

Lähteet: Sintonen & Pekurinen 2006; Weinstein ym. 2009

QALY:n mittaamiseksi tarvitaan tieto hoidon aiheuttamista muutoksista elinvuosissa ja elämänlaadussa (esimerkki 1). Näistä jälkimmäistä voidaan mitata yksittäisiin sairauksiin tai potilasryhmiin keskittyvien (sairausspesifien) tai yleisten (geneeristen) mittareiden avulla. Sairausspesifit mittarit arvioivat tietyn sairauden vaikeusastetta ja vaikutusta elämän laatuun. Tästä esimerkkinä on muun muassa endometriooisia sairastaneille suunnattu elämänlaadun kyselylomake. Geneeriset mittarit puolestaan, kuten esimerkiksi 15D-elämänlaatu-mittari ja EQ-5D, kuvaavat yleisemmin terveydentilan muutoksiin liittyviä elämänlaadun muutoksia (Brooks 1996; Sintonen 2001).

Terveydenhuollossa kehitetyt mittarit eivät kuitenkaan sovellu suoraan sosiaalipalvelujen vaikuttavuuden arviointiin, sillä ne keskittyvät terveydentilassa tapahtuviin muutoksiin. Sen vuoksi sosiaalihuollossa tarvitaan myös omia hyvinvointia mittaavia mittareita. Sosiaalihuollon vaikuttavuuden arviointiin sopivia mittareita ovat esimerkiksi geneeriset arviot elämän laadusta, kuten WHO:n kehittämä elämän laadun mittari (WHOQOL, WHO *Quality of Life*). Lisäksi on olemassa laaja joukko spesifejä mittareita, joita on kehitetty erilaisiin tarpeisiin. Aikuissosiaalityön mittareita ovat esimerkiksi ASCOT-mittari (Netten, Burge, Malley, Potoglou, Towers, Brazier, Flynn, Forder & Wall 2012) ja Avain-mittari. Työelämän tarpeisiin vastaavat Työkyky-mittarit ja vanhuspalveluissa käytetään muun muassa RAI- ja RAVA-mittareita. Näistä osassa asiakas arvioi itse omaa toimintakykyään, kuten Työkyky-mittarin tapauksessa, kun taas osa perustuu ammattilaisen tekemään arvioon potilaan tilanteesta, kuten vanhustyöhön kehitetyt RAI ja RAVA-mittarit. Vaikuttavuuden mittaamiseen onkin olemassa useita tulostittareita, minä vuoksi arviointitutkimuksen tekijät joutuvat pohtimaan, mitkä mittareista soveltuvat parhaiten juuri kyseiseen tutkimukseen. Erilaisia toimintakyvyn ja palvelujen vaikuttavuuden arviointiin kehitettyjä mittareita on esitelty tarkemmin esimerkiksi Terveyden ja hyvinvoinnin laitoksen hallinnoimassa TOIMIA-tietokannassa (2020).

Käytännön vaikuttavuustutkimuksissa itseraportointiin pohjautuvia tulostittareita ei kuitenkaan ole käytetty yhtä paljoa kuin erilaisia rekisteritietoihin pohjautuvia tulostittareita. Yhtäältä tämä on voinut johtua mittareiden puutteesta, mihin on viime vuosina pyritty vaikuttamaan kehittämällä ja validoimalla uusia mittareita. Toisaalta taustalla on ollut myös se, että itsearviointien kerääminen on työllästä eikä tutkimuksissa aina saavuteta riittävää vastausaktiivisuutta, mikä heikentää tulosten luotettavuutta.

Rekisteripohjaiset tulostittarit

Rekisteripohjaisilla tulostittareilla viitataan erilaisiin tietojärjestelmiin ja rekistereihin kirjautuviin tietoihin tai muuttujiin, joiden avulla voidaan arvioida palvelun tai hoidon vaikuttavuutta. Keskeistä tällöin on, että tiedot ovat jälkikäteen kiistattomasti todennettavissa potilas-, asiakastietojärjestelmiin tai rekistereihin tehtävistä merkinnöistä. (Torkki ym. 2017, 38–39.) Tällaisia mittareita ovat esimerkiksi elinikä, syntyneiden lasten lukumäärä, kuolleisuus, sairauseläkkeiden määrä tai huostaanotettujen lasten määrä.

Yksi kansainvälisesti käytetty vaikuttavuuden tulostittari on esimerkiksi ”menetetty toimintakykyiset elinvuodet” (*disability-adjusted life years, DALY*). Se saadaan laskemalla ennenaikaisista kuolemista aiheutuvat elinvuosien menetykset ja lisäämällä näin saatuun lukuun niiden vuosien määrä, kuinka pitkään tautia potevat henkilöt keskimäärin kärsivät sairaudesta (=alentuneen toimintakyvyn vuoksi menetetty vuodet). DALY-arvojen seuraaminen mahdollistaa esimerkiksi väestön terveydentilan seuraamisen paremmin kuin jos tarkasteltaisiin ainoastaan elinaikaa.

Sosiaali- ja terveydenhuollon palvelujärjestelmän laadun ja vaikuttavuuden mittaaminen

Suomessa palvelujärjestelmätason vaikuttavuutta on seurattu useimmiten rekisteripohjaisten tulostittareiden avulla, jolloin on tarkasteltu esimerkiksi menetettyjä elinvuosia, kuolleisuutta tai laitoksissa asuvien määrää. Koska itseraportoituja tulostittareita ei ole juuriikaan systemaattisesti hyödynnetty kansallisella tasolla, on kokonaiskäsitys koko palvelujärjestelmätason vaikuttavuudesta jäänyt vaja-vaiseksi. Torkki kumppaneineen (2017) nostaakin esille niin sanotun *Triple Aim* -viitekehyksen, jonka moniulotteisilla tulostittauksilla päästäisiin tarkastelemaan palvelujärjestelmän vaikuttavuutta kokonaisvaltaisemmin (ks. esim. Valentijn 2015). Tällöin sosiaali- ja

terveyspalvelujen suorituskyvyn arvioinnissa keskiöön nostettaisiin palveluiden kustannukset, väestön terveys (vaikuttavuus) sekä asiakaskokemukset (palvelun laatu). Ajatuksena on, että olemassa olevia mittareita yhdistämällä voitaisiin tulevaisuudessa seurata sekä kokonaiskustannuksia että sosiaali- ja terveyspalvelujen vaikuttavuutta (ks. lisää Torkki ym. 2017, 15–17, 51–61, mm. mittarin pilotoinnista Eksoten aineistossa).

Triple Aim -viitekehyksen ohella viime vuosina on keskusteltu niin sanotusta kansallisesta laaturekisteristä, jonka keskeisenä tavoitteena on seurata palvelujen laatua. Useimmiten tämä on tarkoittanut tulostittareita, mutta mikäli mahdollista myös tutkimusnäyttöön pohjautuvia rakenne- ja prosessimittareita (vrt. *Triple Aim*), jolloin palvelujen vaikuttavuutta voitaisiin seurata paremmin. Lisäksi laaturekistereiden avulla on tarkoitus vertailla eri palveluntuottajien suorituskykyä hoito- tai palveluprosesseissa ja tulosten saavuttamisessa (*benchmarking*, Jonsson ym. 2019, 24).

Toisaalta keskittyminen rekisteripohjaisiin tulostittareihin, kuten kuolleisuuteen, on merkinnyt myös sitä, ettei varsinaiseen vaikuttavuustietoon ole päästy aukottomasti kiinni, sillä kuolleisuuden vaikuttavat myös monet palvelujärjestelmän ulkopuoliset tekijät (ks. Jonsson ym. 2019, 24). Vaikka luotettavat tulostittarit ovat tärkeitä palvelujen vaikuttavuuden tutkimisessa, ei sosiaali- ja terveydenhuollon palvelujärjestelmän suorituskyvyn mittaaminen kansallisilla laaturekistereillä ole siis välttämättä yksistään riittävä edellytys vaikuttavuuden todentamiseen. Ei siis riitä, että pystytään osoittamaan luotettavilla mittareilla, että hoito tai palvelu on laadukasta ja tavoitellut vaikutukset on saavutettu. Vaikuttavuuden todentamiseksi on myös tutkittava sitä, missä määrin mittauksin todetut vaikutukset olisivat nimenomaan tutkimuksen kohteena olevien palvelujen ansiota. (ks. McDavid, Huse & Hawthorn 2013, 16–22.) Esimerkiksi lastensuojelupalvelujen vaikuttavuutta ei voida todeta pelkästään tarkastelemalla kiireellisesti vuoden aikana sijoitettuna olleiden

0–17-vuotiaiden osuutta vastaavan ikäisestä väestöstä³ eli tuottamalla seurantatietoa vaikuttavuudelle asetettujen kriteerien toteutumisesta. Luotettavan tulosmittarin valinnan lisäksi vaikuttavuuden tutkiminen edellyttää tutkimusasetelmaa, jonka avulla on mahdollista todentaa, ovatko palvelut havaitun muutoksen syy. Siirrymme näihin kysymyksiin seuraavaksi.

Kokeellisen tutkimusasetelman perusidea vaikuttavuuden tutkimisessä

Kun arvioidaan sosiaali- ja terveyspalvelujen vaikuttavuutta, arvioidaan sitä, tapahtuiko palveluja käyttäville potilaille tai asiakkaille tavoitteiden mukaisia terveyden tai hyvinvoinnin muutoksia. Tämän lisäksi on pystyttävä arvioimaan kausaalisuhteen toteamiseen (ks. Töttö 2004, 120–135) liittyvää kysymystä (ns. *counterfactual*-kysymys)⁴, olisivatko samat vaikutukset syntyneet myös ilman palveluja (Rossi ym. 2004, 267). Olisi siis pystyttävä osoittamaan, että palvelua käyttäneiden asiakkaiden tai potilaiden ryhmässä havaittua tulosmuuttujan muutosta ei esiintyisi tilanteessa, jossa he eivät olisi saaneet tätä palvelua (McDavid ym. 2013, 16–17, 21). Esimerkiksi arvioitaessa, ovatko lapsen käytösoireet vähentyneet palvelujen seurauksena, tulee lapsen tilannetta mitata samalla, luotettavalla tulosmittarilla (esim. SDQ⁵) vähintään kahdesti: ennen ja jälkeen palvelun saamisen. Jos tavoitellut vaikutukset lapsen käyttäytymiseen saavu-

³ Sotkanet indikaattori Kiireellisesti vuoden aikana sijoitettuna olleet 0–17-vuotiaat, % vastaavan ikäisestä väestöstä on yksi lasten hyvinvoinnin kansallisista indikaattoreista.

⁴ Vaikuttavuuden arvioinnin perusongelma on niin sanottu puuttuva havainto-ongelma. Tällä viitataan siihen, ettei tutkimuksissa voida tutkia yhtä aikaa samoja henkilöitä sekä palvelujen käyttäjinä että ei palveluja saaneina (Rossi ym. 2004, 267). Niin sanottu *counterfactual*-päätelyperiaate vaikuttavuuden arvioinnissa tarkoittaa tätä kausaalianalyysin mahdollistavaa vaihtoehtoisen todellisuuden vastaisen tilanteen vakioimista, johon on kehitetty useita tekniikoita (ks. Loi & Rodrigues 2012).

⁵ *Strengths and Difficulties Questionnaire* eli SDQ on suomennettuna Vahvuudet ja vaikeudet -kysely. SDQ soveltuu Toimia tietokannan mukaan hyvin lasten

tettiin, lisäksi on pystyttävä vakioimaan kaikki muut mahdollisesti lapsen käyttäytymiseen vaikuttavat tekijät eli kilpailevat selitykset sille, miksi havaitut muutokset tapahtuivat (ks. Rubin & Babbie 2008, 256). Jos tämä on mahdollista, voidaan palveluja pitää kyseisten muutosten aiheuttajina.

Kun halutaan arvioida palvelujen vaikuttavuutta, tulisi pyrkiä tutkimusasetelmaan, jonka avulla palvelujen vaikuttavuutta voidaan mitata mahdollisimman luotettavasti. Satunnaistettua koeasetelmaa pidetään luotettavimpana tapana tutkia vaikuttavuutta. **Satunnaisesti kontrolloitu koeasetelma** (*randomized controlled trial, RCT*) perustuu nimensä mukaisesti kahden ryhmän eli koe- ja kontrolliryhmän satunnaistamiseen. Asetelman vahvuus on siinä, että onnistuessaan satunnaistaminen vakioi kaikki muut (intervention ohella) tulostuuttujaan vaikuttavat havaitut ja ei-havaitut sekoittavat tekijät. Satunnaistamiseen on olemassa erilaisia menetelmiä, joista voi valita kyseiseen tutkimukseen sopivimman (Craig, Dieppe, Macintyre, Michie, Nazareth & Petticrew 2008, 10–11). Rubin ja Babbie (2008, 253) huomauttavat, että on tärkeää erottaa toisistaan satunnainen ryhmiin arpominen ja satunnaisotanta. Onnistunut satunnaisotanta takaa sen, että aineistosta saadut tulokset ovat yleistettävissä perusjoukkoon tai kohdeväestöön. Asiakkaiden satunnaistaminen koe- ja kontrolliryhmiin puolestaan takaa sen, että koe- ja kontrolliryhmien vastaajat eivät eroa toisistaan muiden muuttujien kuin testattavan palvelun käytön suhteen, ja siten kausaalisten päätelmien tekeminen on luotettavaa. (Rubin & Babbie 2008, 253.)

On olemassa erilaisia koeasetelmia (ks. Rubin & Babbie 2008, 252–253). Yksinkertaisessa satunnaistetussa kokeellisessa tutkimusasetelmassa tutkittavat jaetaan satunnaisesti koe- ja kontrolliryhmään siten, että molemmat ryhmät ovat lähtötilanteessa taustatekijöiltään (esim. sukupuoli, ikä, ammattiasema) samanlaisia. Tutkimuksen kohteena oleva palvelu toteutetaan ainoastaan koeryhmässä ja palve-

psykkisten oireiden tunnistamiseen (ks. <https://www.terveysportti.fi/dtk/tmi/koti>).

lun saamisen jälkeen molemmista ryhmistä mitataan tulosuuttujan arvot ja verrataan niitä lähtötilanteeseen⁶. Jos palvelu lisäisi terveyttä tai parantaisi hyvinvointia, tulisi palvelua saaneiden koeryhmän jäsenten tulosuuttujan arvojen olla erisuuria kuin kontrolliryhmässä, joka ei saanut palvelua. Jos tulosuuttujan arvoissa havaitaan tilastollisesti merkitsevä ero koe- ja kontrolliryhmien välillä, voidaan todeta, että palvelulla oli vaikutusta tulosuuttujaan (palvelun vaikuttavuus, *program effect*⁷). (Robson 2001, 89; Rossi ym. 2004, 242–243.) Esimerkiksi Ihmeelliset vuodet (*The Incredible Years, IY*) intervention vaikuttavuutta lasten käyttäytymisvaikeuksiin mitattiin ECBI⁸-tulomittarin muutosten (6 kk seuranta-lähtötilanne) erotuksena koe- ja kontrolliryhmän välillä, jolloin vaikuttavuus oli 20,33 ECBI yksikön vähennys (ks. O’Neill ym. 2013). Alla on esimerkki yksinkertaisesta satunnaistetusta koeasetelmasta, jossa ei tehdä alaryhmäanalyyskejä (esimerkki 2).

⁶ Satunnaistamisen onnistuessa tulomittarin arvojen tulisi olla lähtötilanteessa samat koe- ja kontrolliryhmässä (ks. esim. O’Neill, McGilloway, Donnelly, Bywater & Kelly 2013).

⁷ Lisäksi käytetään termejä *program’s impact* tai *outcome change* (ks. esim. Rossi ym. 2004, 232).

⁸ Eyberg Child Behavior Inventory (ECBI)

Esimerkki 2. Satunnaistettu kokeellinen tutkimusasetelma

		Tilanne ennen palvelun antamista		Tilanne palvelun saamisen jälkeen
Koeryhmä	R	O ₁	X	O ₂
Kontrolliryhmä	R	O ₁		O ₂

- R = Tutkittavat arvotaan ensin koe- tai kontrolliryhmään satunnaisesti esim. satunnaislukutaulukon avulla
- O₁ = Alkumittaus ja O₂ = Loppumittaus/Kontrolliryhmän mittaus vastaavina ajankohtina
- X = Palvelu, jonka vaikuttavuutta tutkitaan; kontrolliryhmälle ei anneta tutkittavaa palvelua

Lähteet: Mukailleen Robson (2001, 89) ja Rubin & Babbie (2008, 251)

Koeasetelma on suunniteltava ja toteutettava huolellisesti siten, että kaikki kontrolloitavissa olevat harhalähteet saadaan minimoitua.⁹ Ideaalitulanteessa satunnaistaminen tulisi toteuttaa niin sanotusti sokkona. **Sokkouttaminen** (*blinding*) tarkoittaa sitä, että potilas tai asiakas, hänelle hoitoa tai palveluja antava ammattilainen tai tutkija itsekään ei tiedä, kuka kuuluu koe- tai kontrolliryhmään. Ensinnäkin potilaiden tai asiakkaiden tutkimusryhmiin jakaminen tulee pitää salassa (*allocation concealment*) eli tutkittavien rekrytoinnista vastaava tutkija ei saa itse osallistua satunnaistamismenettelyyn. Jos potilas, asiakas tai hoitohenkilökuntakaan eivät tiedä tutkittavien jäsenyyttä, puhutaan kaksoissokkouttamisesta (*double-blind*). (Komalainen, Vuorela & Malmivaara 2014, 1440.) Seuraavaksi kuvaamme esimerkin 3 avulla onnistunutta satunnaistetun koeasetelman

⁹ Esimerkiksi Joanna Briggs Instituutti on määritellyt tarkistuslistan kokeellisten tutkimusasetelmien ”metodologisen laadun arviointiin ja tutkimuksen tuloksiin vaikuttavan harhan riskin tunnistamiseen” (ks. <https://www.hotus.fi/jbin-kriittisen-arvioinnin-tarkistuslistat/>). Esimerkiksi tätä tarkistuslistaa voi käyttää myös satunnaistetun kokeellisen tutkimusasetelman suunnittelun ja toteutuksen muistilistana. Tarkistuslista käsittää 10 arviointikriteeriä, joista laadukkaasti toteutetussa kokeellisessa tutkimuksessa mahdollisimman moni täyttyy.

käyttöä. Esimerkki perustuu niin kutsuttuun Finger-tutkimukseen (Ngandu, Lehtisalo, Solomon, Levälahti, Ahtiluoto, Antikainen, Bäckman, Hänninen, Jula, Laatikainen, Lindström, Mangialasche, Paajanen, Pajala, Peltonen, Rauramaa, Stigsdotter-Neely, Strandberg, Tuomilehto, Soininen & Kivipelto 2015).

Esimerkki 3. FINGER-tutkimus esimerkkinä onnistuneesta satunnaistetusta koeasetelmasta

FINGER-tutkimuksessa (Ngandu ym. 2015) tutkittiin elintapojen muutoksiin tähtäävän intervention kognitiivisia vaikutuksia. Tutkimus toteutettiin Helsingissä, Vantaalla, Kuopiossa, Seinäjoella, Oulussa ja Turussa. Monitahoisella interventiolla pyrittiin vaikuttamaan ikääntyneiden ruokavalioon, liikuntaan, kognitiivisiin ja sosiaalisiin toimintoihin sekä sydän- ja verisuonisairauksien riskien hallintaan.

Tutkimuksen kohderyhmänä olivat 60–77 -vuotiaat henkilöt, joilla oli kohonnut muistisairauden riski. Tutkittavat henkilöt rekrytoitiin syyskuun 2009 ja marraskuun 2011 välisenä aikana. Tutkimukseen suostuneista 631 henkilöä satunnaistettiin koeryhmään ja 629 henkilöä kontrolliryhmään. Kontrolliryhmään osallistuvat henkilöt saivat normaalikäytännön mukaista terveysneuvontaa. Koeryhmän jäsenille tarjottiin lisäksi moniulotteinen interventio, jolla pyrittiin vaikuttamaan ikääntyvien henkilöiden ruokailutottumuksiin, liikuntaan, muistiin ja sydän- ja verisuonisairauksien riskien hallintaan. Tutkimuksessa sovellettiin kaksoisokkoutuksen periaatetta ja satunnaistaminen toteutettiin lohkoittain (*blocked randomization*). Ensimmäisenä tulostuutujana käytettiin kognitiivista suorituskykyä mittaavaa normalisoitua NTB indeksiä.

Tulokset osoittivat, että kognitio parani noin 25 prosenttia enemmän interventioyhmässä kuin kontrolliryhmässä (taulukko E1.1) ja että tämä ero ryhmien välillä oli tilastollisesti merkitsevä (Ngandu ym. 2015). FINGER oli ensimmäisiä tutkimuksia, jossa pystyttiin todentamaan elintapoihin vaikuttavan intervention vaikuttavuus ikääntyvillä riskiryhmiin kuuluvilla henkilöillä.

Taulukko E1.1 FINGER-tutkimuksen päätulokset

	Kahden vuoden seuranta
NTB indeksin muutos koeryhmässä	0.20
NTB indeksin muutos kontrolliryhmässä	0.16
Muutosten erotus koe- ja kontrolliryhmän välillä per vuosi	0.02

Lähde: Ngandu ym. 2015

Satunnaistettua kokeellista asetelmaa suunniteltaessa kannattaa varautua erityisesti seuraaviin asioihin. Satunnaistaminen voi joskus epäonnistua, jolloin koe- ja vertailuryhmien välille voi muodostua systemaattinen ero joissakin tulosmuuttujaan vaikuttavissa tekijöissä. Koeryhmään voi esimerkiksi valikoitua enemmän naisia kuin miehiä. Jos sukupuoli vaikuttaa tulosmuuttujaan, on tässä tapauksessa erheellistä päätellä, että havaitut muutokset tulosmuuttujassa johtuisivat pelkästään tutkittavasta palvelusta tai toteutetuista toimenpiteistä. Tällaista valikoitumisharhaa voitaisiin korjata vakioimalla sukupuoli analyysseissa. Vaikka satunnaistaminen näennäisesti vähentääkin tehtävien mittausten tarvetta, myös alaryhmäanalyysija (esim. selittääkö ikä vaikuttavuutta) sisältävissä tutkimuksissa on joka tapauksessa mitattava ja kontrolloitava väliin tulevia tekijöitä (Vuorela ym. 2014, 1546). Satunnaistamiseen liittyvien ongelmalanteiden lisäksi keskeisiä satunnaistettuun koeasetelmaan liittyviä harhamahdollisuuden lähteitä ovat esimerkiksi sokkouttamisen epäonnistuminen ja **tutkittavien vähittäinen poistuminen seurannasta** (*attrition*) (esim. Komulainen ym. 2014)¹⁰.

Esimerkiksi tietäntyyppisissä leikkaustoimenpiteissä (Williams, Burden-Teh & Nunn 2015) ja erityisesti sosiaalialalla sokkouttamisen toteuttaminen voi olla haasteellista. Sosiaalisten interventioiden toteuttamista, kuten esimerkiksi palveluohjauksen antamista tai lasten sijoituspäätöksiä, voi olla mahdotonta salata asiakkailta ja henkilökunnalta samalla tavalla kuin esimerkiksi lääketutkimuksessa tutkittavan lääkkeen ja plasebo-tabletin suhteen voidaan menetellä. Tulostittauksen arvioinnin sokkouttaminen on helpompi toteuttaa, mutta Komulainen ja tutkimusryhmä (2014, 1440) arvioivat kuitenkin, että sokkouttamisen puuttuminen on vakavinta silloin, jos käytetään asiakkaan itsearviointiin perustuvia tulostittareita: sen

¹⁰ Sisäisen validiteetin harhatekijöitä käsitellään tarkemmin artikkelin luvussa ”Vaikuttavuuden tutkiminen havainnoivilla tutkimusasetelmilla. Havainnoivat tutkimusasetelmat ja niihin liittyvät harhamahdollisuudet”.

sijaan jos tulostittari on objektiivinen (esim. painoa mittaava vaakana), ei tutkijan sokkouttamisen puute ole niin suuri epäkohta.

Vaikka satunnaistaminen olisikin onnistunut ja koe- ja kontrolliryhmät olisivat lähtötilanteessa täysin samanlaisia, voi systemaattista eroa joissakin keskeisissä taustatekijöissä muodostua tutkimuksen aikana myös tutkittavien vähittäisen poistuman kautta. Tällainen tilanne syntyy esimerkiksi silloin, jos kontrolliryhmään osallistuvista perheellisistä osa lopettaa tutkimuksen kesken, koska he kokevat, että tutkimus vie liikaa aikaa eikä hyödytä heitä riittävästi. Jos vastaavaa poistumaa ei tapahdu koeryhmän kohdalla, ja ero on tilastollisesti merkitsevä, perheeseen liittyvä valikoituminen voi vääristää tuloksia (valikoitumisharha, *selection bias*). **Valikoitumisharhaa** muodostuu siis, jos koe- ja vertailuryhmät eroavat toisistaan systemaattisesti jonkin tekijän suhteen joko tutkimuksen alussa tai sen aikana ja näillä tekijöillä on yhteys tulosmuuttujaan. (Rossi ym. 2004, 270–271.) Valikoitumisharhaa voi korjata tilastollisilla menetelmillä¹¹, joten olennaisimpia tulosmuuttujaan vaikuttavia asioita on suotavaa mitata, vaikka käytettäisiin satunnaistettua koeasetelmaa.

Harhojen minimoimiseksi satunnaistetuissa kokeellisissa tutkimuksissa on tapana myös toteuttaa **protokollan mukaisen analyysin** (*per-protocol analysis*) sijaan niin sanottu **hoitoaikeen mukainen analyysi** (*intention to treat analysis, ITT*) (ks. McCoy 2017). Eli sen sijaan, että analyysi rajattaisiin vain palveluja tosiasiaassa saaneisiin, kaikki koe- ja vertailuryhmiin satunnaistetut ihmiset analysoidaan siinä ryhmässä, johon heidät on satunnaistettu – vaikka he vaihtaisivatkin ryhmää tai jättäytyisivät pois kesken tutkimuksen. Tällä tavoin asetelma on lähempänä luonnollista tilannetta, jossa hyppäyksiä eri ryhmien välillä ja hoidon keskeyttämistä voi tapahtua. Toisaalta

¹¹ Katso lisää kontrollimuuttujien käyttämisestä ja sekoittavien tekijöiden tilastollisesta vakioinnista luvusta ”Epävarmuustekijöiden kontrolloiminen tilastollisten analyysimenetelmien avulla”.

se tarkoittaa myös sitä, että seurannasta pois jääneiden puuttuvat vastaukset on korvattava (imputoitava) jollain tavalla¹². (McCoy 2017.)

Tutkimustulosten raportoinnin ja soveltamisen näkökulmasta on tärkeää tiedostaa, että satunnaistetun kokeellisen tutkimusasetelman **ulkoinen validiteetti** (*external validity*) eli tulosten yleistettävyys vastaaviin olosuhteisiin voi kärsiä **sisäisen validiteetin** (*internal validity*) eli vaikuttavuutta koskevien päätelmien luotettavuuden korostamisen kustannuksella. Syynä voi olla esimerkiksi se, että koeasetelman käytännön toteutus on kallista ja tutkimusaineiston optimaalinen koko on laskettava tarkkaan vähimmäisrajalle tilastollisten voimalaskelmien avulla. Satunnaistetussa kontrolloidussa koeasetelmassa määritellään usein myös tiukat rekrytointikriteerit, ja hoidon toteutus on tiukasti ohjattua ja valvottua, minkä takia aineisto ei välttämättä vastaa hoidon toteutumista sosiaali- ja terveydenhuollon normaalikäytännöissä (Rannanheimo & Jauhonen 2018, 20). Esimerkiksi vakavien psykiatristen häiriötilojen kriteerit täyttävät henkilöt voidaan helposti sulkea pois tutkimuksista, vaikka he tositilanteissa ovatkin palvelujen käyttäjiä. Siten satunnaistetun koeasetelman tutkimustuloksia voi olla haasteellista yleistää käytännön tilanteisiin tai eri konteksteihin, kuin missä tutkimus on tehty.

Ideaalitilanteessa sekä tutkimuksen ulkoinen että sisäinen validiteetti ovat mahdollisimman hyviä. Silloin myös satunnaistettujen koeasetelmien tulokset voitaisiin yleistää tutkimusten perusjoukkoihin. Tähän asti olemme käsitelleet satunnaistettua koeasetelmaa niin kutsutun satunnaistetun kontrolloidun koeasetelman (ns. *RCT*; *efficacy trial*; *explanatory clinical trial*) näkökulmasta. Satunnaistettuja, kontrolloituja koeasetelmia voidaan toteuttaa myös arkivaikeuttavuuden tutkimiseen normaaleissa sosiaali- ja terveydenhuollon toimintaympäristöissä (ns. *effectiveness trial*; *pragmatic clinical trial*). Satunnaistettu koeasetelma sijoittuu usein johonkin kohtaan näiden

¹² Puuttuvien tietojen korvaamiseen on olemassa erilaisia menetelmiä. Esimerkiksi tässä artikkelissa esitellyssä Nuorten osa-aikatyö tutkimuksessa (esimerkki 4) on hyödynnetty moni-imputointia, josta voi lukea tarkemmin kyseisestä artikkelista (Monahan, Lee & Steinberg 2011).

ääripäiden välille¹³. (Sackett 2011.) Niin kutsutussa **käytännöllisorientoituneessa satunnaistetussa koeasetelmassa** arvioidaan sosiaali- ja terveydenhuollon normaalihenkilöstön tavanomaisille asiakkaille tai potilaille tavanomaisessa hoito- ja palveluympäristössä toteuttamia hoitoja tai palveluja tarkoituksenmukaisen vertailuryhmän ja kliinisesti soveltuvan tulostamituksen avulla. Koska arviointi ei kohdistu pelkästään kliinisten asiantuntijoiden tiukasti rajatuille osallistujille toteuttamiin kontrolloituihin hoitoihin, käytännöllinen satunnaistettu koeasetelma tuottaa RCT-asetelmaa tarkemmin sosiaali- ja terveydenhuollon normaalikäytäntöön yleistettävää ja sovellettavaa tietoa (parempi ulkoinen validiteetti). (ks. Sackett 2011; Williams ym. 2015; Gamerman, Cai & Elsässer 2018.)

Satunnaistettu kokeellinen tutkimusasetelma takaa onnistuessaan luotettavat päätelmät nimenomaan siitä, voidaanko palvelun käyttäminen yhdistää tutkimuksessa havaittuihin vaikutuksiin. Kun näitä tutkimustuloksia tulkitaan, on lisäksi tärkeää tarkastella sitä, missä määrin tutkimuksen otos vastaa tutkimuksen kohdeväestöä ja ratkaista tämän perusteella tutkimustulosten yleistettävyyttä käytännön tilanteisiin. Koska satunnaistettujen, kokeellisten tutkimusasetelmien aineistot ovat usein pieniä, on lisäksi hyvä tarkastaa, onko olemassa systemaattisia kirjallisuuskatsauksia, joissa olisi tehty meta-analyysi kyseisten palvelujen vaikuttavuudesta (Malmivaara & Komulainen 2014).

¹³ Vertaa ehkäisevien palvelujen tutkimussykli: palvelujen tehon, vaikuttavuuden ja laaja-alaisen levittämisen mahdollisuuksien vaiheittainen tutkiminen (ks. Gottfredson ym. 2015).

- **Koeryhmä (treatment, intervention tai experimental group):** Ryhmä, jossa toteutetaan interventio (hoito/palvelu) ja jonka saamia tulomuuttujan arvoja verrataan kontrolliryhmään.
- **Kontrolliryhmä (control group):** Ryhmä, joka ei osallistu interventioon (saa esim. lumelääkettä, osallistuu ”vanhaan” normaalikäytännön mukaiseen palveluun tai on ”odotuslistalla”) ja jonka saamia tulomuuttujan arvoja verrataan koeryhmän saamiin tulomuuttujan arvoihin.
- **Satunnaistaminen (randomization):** Menettely, jonka perusteella perusjoukkoon kuuluvat jäsenet jaetaan satunnaisesti koe- ja kontrolliryhmään siten, että jokaisella perusjoukkoon kuuluvalla on sama todennäköisyys tulla valituksi kumpaan tahansa ryhmistä. Satunnaistaminen voidaan toteuttaa esimerkiksi nopanheiton tai satunnaislukugeneraattorien avulla.
- **Satunnaistettu kontrolloitu koeasetelma (randomized controlled trial, RCT):** Tutkimusasetelma, jossa koe- ja kontrolliryhmät muodostetaan satunnaistamisen avulla.
- **Palvelun tai hoidon vaikuttavuus (program effect tai program’s impact tai outcome change):** Toiminnan vaikutuksesta tapahtuva muutos toiminnan perimmäisessä tavoitteessa eli se osuus tulomuuttujan muutoksesta, joka on käytetyn palvelun tai hoidon aiheuttamaa, kun kaikki muut tulomuuttujan vaikuttavat tekijät on kontrolloitu.
- **Ulkoinen validiteetti (external validity):** Tulosten yleistettävyys tutkitun kohdejoukon ulkopuolelle. Eli kyse on siitä, voidaanko tutkimustuloksia soveltaa muissa vastaavissa sosiaali- ja terveydenhuollon toimintaympäristöissä.
- **Sisäinen validiteetti (internal validity):** Tulosten luotettavuus eli arvio siitä, kuinka suurella varmuudella tavoiteltu muutos on nimenomaan annetun menetelmän, palvelun tai politiikkatoimen aiheuttamaa ja millaisia harhamahdollisuuksia vaikuttavuutta koskeviin päätelmiin sisältyy.

Vaikuttavuuden tutkiminen havainnoivilla tutkimusasetelmilla

Havainnoivat tutkimusasetelmat ja niihin
liittyvät harhamahdollisuudet

On tilanteita, joissa satunnaistettu kokeellinen tutkimusasetelma ei ole paras mahdollinen valinta. Luotettavan kokeellisen aineiston kerääminen voi olla kallista ja siksi aineistot voivat jäädä pieniksi. Eri-tyisesti RCT-asetelmissa ei niille tyypillisen lyhyen seuranta-ajan (ks. Williams ym. 2015) vuoksi voida välttämättä luotettavasti havaita sellaisia tapahtumia, jotka ovat hyvin harvinaisia tai joiden ilmaantumiseen kuuluu oletettavasti pitkä aika. Joissakin tilanteissa satunnaistetun koeasetelman soveltaminen ei ole mahdollista eettisistä syistä. Esimerkiksi jos tiedettäisiin, että uusi palvelu on vaikuttavampi kuin nykyinen palvelu, olisi epäeettistä evätä uuden palvelun saanti kontrolliryhmään kuuluvilta henkilöiltä. Usein tällaista etukäteistietoa ei kuitenkaan ole, jolloin nimenomaan on eettisesti perusteltua testata, onko uusi palvelu tai hoitomuoto nykyistä hoitokäytäntöä vaikuttavampi.

Sosiaali- ja terveyspalvelujen vaikuttavuutta voidaan tutkia myös niin kutsutuilla **havainnoivilla tutkimusasetelmilla** silloin, kun kokeellisten tutkimusasetelmien käyttäminen ei ole eettisistä tai muista syistä johtuen mahdollista. Havainnoivat tutkimusasetelmat ovat ei-kokeellisia tutkimusasetelmia, joissa tutkija ei vaikuta (*manipulate*) tutkittavien sijoittumiseen **koe- ja vertailuryhmään** (Kirk 2013, 6). Havainnoivien asetelmien etuna on mahdollisuus kerätä suurempi aineisto, jolloin on mahdollista havaita hienovaraisempia yhteyksiä kuin satunnaistettujen kokeellisten tutkimusten usein pienillä aineistoilla. Havainnoivissa asetelmissa tutkimustulosten yleistettävyyys (ulkoinen validiteetti) on yleensä parempi, koska tutkimukseen hyväksytään mukaan toisistaan voimakkaammin eroavia ihmisiä (heterogeenisempi aineisto) kuin kokeellisissa tutkimusase-

telmissä. Havainnoiviin tutkimusasetelmiin kuuluvat puolikokeellinen (*quasi-experimental*) (mm. Rossi ym. 2004, 237) ja esikokeellinen (*pre-experimental*) tutkimusasetelma (Rubin & Babbie 2008, 255–264).

Kun vaikuttavuutta arvioidaan jollakin havainnoivalla tutkimusasetelmalla, on vaikuttavuuden arvioinnin perusongelma eli kausaalisuhteen päättämiseen liittyvä todellisuudenvastainen kysymys siitä, olisivatko samat vaikutukset syntyneet myös ilman palveluja (Rossi ym. 2004, 267), ratkaistava jollain muulla tavalla kuin satunnaistamismenettelyllä. **Puolikokeellisessa tutkimusasetelmassa** jäljitelläänkin mahdollisimman paljon kokeellista tutkimusasetelmaa. Asetelmaan kuuluu koe- ja vertailuryhmät, mutta tutkittavia ei satunnaisteta näihin ryhmiin. Kontrolliryhmän sijaan puhutaan usein vertailuryhmästä, ja puolikokeellista tutkimusasetelmaa nimitetään ”**ei-vastaavien vertailuryhmien asetelmaksi**” (*nonequivalent comparison group design*) (Rubin & Babbie 2008, 255). Valikoitumisharhan minimoimiseksi on pyrittävä erityisesti varmistamaan se, että koe- ja vertailuryhmät olisivat lähtötilanteessa mahdollisimman samankaltaiset. Vertailtaviin ryhmiin voidaan esimerkiksi valita mahdollisimman samanlaisia toimipaikkoja tai yksilöitä. Otantapopulaation osituksen keinoin kohdeväestöä voidaan jakaa ryhmiin ja varmistaa, että vertailtaviin ryhmiin tulee yhtä paljon esimerkiksi miehiä ja naisia tai eri ikäryhmiin kuuluvia henkilöitä. Luotettavimmat päätelmät vaikuttavuudesta saavutetaan, jos tutkimusasetelma sisältää sekä mahdollisimman hyvin koeryhmää vastaavan vertailuryhmän että aikasarja-aineiston (Rossi ym. 2004; Rubin & Babbie 2008).

Puolikokeellinen tutkimusasetelma voidaan muodostaa joko takautuvasti tai reaaliaikaisesti etenevänä. Vastaavasti Vuorela työryhmineen (2014) määrittelee puolikokeellisten (havainnoivien) tutkimusasetelmien päätyypeiksi **altistuslähtöisen** ja **tapauslähtöisen** tutkimusasetelman. Altistuslähtöisellä tutkimusasetelmalla he tarkoittavat etenevää **kohorttitutkimusasetelmaa** (*cohort study*) ja

tapauslähtöisellä tutkimusasetelmalla takautuvasti toteutettua, niin kutsuttua tapaus-verrokkitutkimusta (case-control study). Reaaliaikaisesti etenevässä puolikokeellisessa tutkimusasetelmassa edetään satunnaistamista lukuun ottamatta muuten mahdollisimman vastaavasti kuin kokeellisessa tutkimusasetelmassa (ks. Rubin & Babbie 2008, 255–256). Hyvä lähtökohta reaaliaikaisesti etenevälle puolikokeelliselle tutkimusasetelmalle on, jos voidaan paikantaa kaksi toisilleen verrannollista ryhmää, joista vain toiselle annetaan tutkittava hoito tai palvelu (esim. Saarni 2009). Sekä koe- että vertailuryhmissä toteutetaan lähtömittaus, minkä jälkeen vain koeryhmä saa hoitoa tai käyttää tutkittavaa palvelua. Sopivan seuranta-ajan kuluttua mitataan molempien ryhmien tilanne samoilla, mahdollisimman luotettavilla tulosmittareilla. Jos tulosmuuttujan muutos koe- ja vertailuryhmien välillä on odotusten mukainen, se voidaan puolikokeellisessa tutkimusasetelmassa tulkita palvelun vaikuttavuudeksi sillä edellytyksellä, että ryhmiin valikoituminen ja tulosmuuttujaan yhteydessä olevat muut mahdolliset sekoittavat tekijät saadaan onnistuneesti kontrolloitua. (Rubin & Babbie 2008, 255–256.)

Puolikokeellinen tutkimusasetelma voidaan muodostaa myös takautuvasti rekisteriaineistojen¹⁴ avulla (ns. *ex post facto*). Takautuvan tutkimuksen lähestymistapa on tyypillinen epidemiologisessa tutkimuksessa. Tapaus-verrokki-asetelmassa toisiinsa vertailtavat ryhmät muodostetaan takautuvasti määritellyn asiantilan toteutumisen suhteen. Tutkittavat henkilöt voidaan jakaa tapauksiin ja verrokkeihin esimerkiksi sen mukaan, ovatko he sairastuneet Alzheimerin tautiin (tapaukset) vaiko eivät (verrokkit). Tapaus-verrokki-tutkimuksessa etsitään niitä tekijöitä, jotka selittäisivät eroja tulosmuuttujien arvoissa tapaus- ja verrokkiryhmien välillä. Toinen mahdollisuus on hyödyntää esimerkiksi asiakastietorekisteriin kertynyttä tietoa ja rakentaa niin sanottu *catch-up*-tutkimusasetelma. Tämä on mahdollista silloin, jos on olemassa jo kerättyä aineistoa tietyistä tutkimuk-

¹⁴ Katso lisää rekisteriaineistojen käsittelystä ja käyttämisestä esim. Sund (2003) ja Sund, Nylander ja Palonen (2004).

sen kannalta kiinnostavista osaryhmistä, joita voidaan seurata ajassa eteenpäin ja ottaa heihin myöhemmin yhteys. Tällaisessa asetelmassa tarvitaan vain yksi yhteydenotto, koska aineistossa on jo valmiiksi tietoa edeltävästä, tutkimuksellisesti kiinnostavasta asiantilasta, kuten onko kyseinen henkilö käyttänyt tiettyä palvelua lapsuudessaan vai ei. (ks. Rutter 1994.)

Kausaalinäytön asteelta puolikokeellista tutkimusasetelmaa heikompa versiota edustavat **esikokeelliset tutkimusasetelmat**. Tällaisia tutkimusasetelmia ovat yhden ryhmän **ennen-jälkeen-tutkimusasetelma** (*pre-post design; before and after design*) ja **aikasarja-asetelma** (*time series design*). Esikokeellisissa tutkimusasetelmissä käytetään kausaalipäättelyn välineenä vertailuryhmän sijaan niin sanottua **refleksiivistä kontrollia** (*reflexive control(s)*) eli vertailua aiempaan tilanteeseen (Rossi ym. 2004, 289–295). Ennen-jälkeen-tutkimusasetelmassa verrataan toisiinsa ennen palvelun toteuttamista annettua alkumittausta ja palvelun toteuttamisen jälkeen tehtyä jälkimittausta. Aikasarja-asetelman käyttäminen eli sekä ennen ja jälkeen palvelun saamisen toteutetut useat tulosmuuttujan mittaukset vahvistavat yleensä esikokeellista tutkimusasetelmaa, sillä näin voidaan jäsentää mahdollista muutosta palvelun saamista edeltävässä ja sen jälkeisessä trendissä (esim. Rossi ym. 2004, 290–295).

Esikokeellisiin tutkimusasetelmiin luokitellaan myös **poikkileikkaustutkimusasetelmat**, joihin kuuluu esimerkiksi kertaluontoinen yhteen tutkimusryhmään kohdentuva tapaustutkimus ja lomakekysely. Rubin ja Babbien (2008) mukaan tutkimuksen sisäinen validiteetti voidaan turvata parhaiten, jos tutkimuksessa on käytössä sekä koe- ja vertailuryhmäasetelma että aikasarja-aineisto. Täten vaikuttavuuden tutkimisen näkökulmasta heikoimpia tutkimusasetelmia ovat kertaluontoiset, yhteen ryhmään kohdentuvat tutkimusasetelmat. Poikkileikkausaineistolla ei voi todistaa, onko jokin asiantila muuttunut vai ei tai onko havaittujen ilmiöiden välillä kausaalista yhteyttä. Poikkileikkausaineistoilla voidaan kuitenkin erilaisia tilastollisia tekniikoita hyödyntämällä testata vaikuttavuutta koskevia

hypoteeseja. Hypoteesien testaamisen edellytyksenä on, että täsmennetään etukäteen oletukset siitä, kuinka palvelu tuottaa toivotun muutoksen ja muodostetaan tarkat, testattavat hypoteesit ennen tilastollisten analyysien suorittamista. (Rutter 1994.)

Vaikuttavuuden tutkimisen haasteena on se, että palvelujen vaikuttavuutta voivat ehdollistaa monenlaiset tuotetusta palvelusta riippumattomat tekijät (mm. Torkki ym. 2017, 35–36). Kaikkia sekoittavia tekijöitä ei useinkaan tunneta, minkä vuoksi niiden kontrollointi havainnoivissa tutkimuksissa voi jäädä puutteelliseksi. Tutkimustulosten raportoinnin ja tulosten soveltamisen näkökulmasta onkin tärkeää tiedostaa, että havainnoiviin tutkimusasetelmiin liittyy kokeellisia asetelmia enemmän palvelujen vaikuttavuutta koskevien päätelmien luotettavuutta heikentäviä **harhamahdollisuuksia** (*bias*)¹⁵. Erilaisten harhan lähteiden takia puolikokeellisissa tutkimusasetelmissä vertailtavat ryhmät eivät aina ole vertailukelpoisia, jolloin tutkitun hoidon tai palvelun vaikutuksen estimaatti on joko liian pieni tai liian suuri (Rossi ym. 2004, 267). Myös satunnaistetussa koeasetelmassa koe- ja kontrolliryhmien vertailukelpoisuus voi kärsiä, jos kokeen aikana joko koe- tai kontrolliryhmästä poistuu esimerkiksi tilastollisesti merkitsevästi suurempi osuus nuoria (vähittäinen poistuma, ks. Rossi ym. 2004, 274). Suurimmat epävarmuustekijät liittyvät esikokeellisiin tutkimusasetelmiin.

¹⁵ Korostamme, että tutkimusasetelmasta riippumatta tieteellinen tutkimus on toteutettava joka vaiheessa korkealaatuisesti ja kaikkia hyvän tieteellisen käytännön periaatteita noudattaen. Erilaisten harhojen muodostumista voidaan vähentää, kun tutkimus toteutetaan laadukkaasti ja huolellisesti. Myös erilaisten havainnoivien tutkimusasetelmien suunnittelun ja toteutuksen apuna voidaan käyttää mallina esimerkiksi Joanna Briggs Instituutin kirjallisuuskatsauksia varten tekemiä alkuperäistutkimusten kriittisen arvioinnin listoja (ks. <https://www.hotus.fi/jbin-kriittisen-arvioinnin-tarkistuslistat/>).

- **Havainnoiva tutkimusasetelma (observational tai non-experimental study):** Tutkimusasetelma, jossa koe- ja vertailuryhmän muodostaminen ei perustu satunnaistamiseen eikä ole siten tutkijan kontrolloitavissa, kuten kontrolloiduissa kokeellisissa tutkimusasetelmissa.
- **Puolikokeellinen tutkimusasetelma (quasi-experimental study):** Tutkimusasetelma, jossa koe- ja vertailuryhmät eivät perustu satunnaistamiseen, mutta niiden muodostamisessa jäljitellään kokeellista tutkimusasetelmaa mahdollisimman hyvin. Puolikokeelliseen tutkimusasetelmaan voidaan myös päätyä, jos kokeellisen tutkimusasetelman satunnaistaminen epäonnistuu.
- **Esikokeellinen tutkimusasetelma (pre-experimental study):** Tutkimusasetelma, jonka avulla voidaan esitellä hypoteeseja palvelujen vaikuttavuudesta ja vaikuttavuutta ehdollistavista tekijöistä, mutta ei todentaa niitä.
- **Koeryhmä ja vertailuryhmä (treatment group ja comparison group):** Toisin kuin kokeellisissa tutkimusasetelmissa, havainnoivissa tutkimusasetelmissa puhutaan koe- ja vertailuryhmästä. Tällä painotetaan, ettei ryhmien muodostaminen perustu satunnaistamiseen, kuten satunnaistetussa kokeellisessa tutkimusasetelmassa. Silti koeryhmä edustaa ryhmää, jossa toteutetaan interventio (esim. saa/sai tiettyä palvelua), kun taas vertailuryhmä edustaa nimensä mukaisesti koeryhmään verrattavaa joukkoa, joka ei osallistu(nut) tutkittavaan interventioon.
- **Harha (bias):** Arvio (estimaatti) palvelun vaikuttavuudesta on liian suuri tai liian pieni johtuen koe- ja vertailuryhmään liittyvästä valikoitumisesta tai mallin ulkopuolisista muuttujista.

Valikoitumisharha on yksi keskeisistä vaikuttavuustutkimuksen harhan lähteistä silloin, kun käytetään havainnoivia tutkimusasetelmia (ks. esim. Rossi ym. 2004, 268–271). Tällöin koe- ja vertailuryhmiin valikoitumista ovat aiheuttamassa tekijät, jotka ovat yhteydessä tulosuuttajaan. Valikoitumisharha on todennäköinen esimerkiksi silloin, kun sosiaali- ja terveyspalvelujen käyttäjiksi valikoituvat korkean terveysriskin tai suuren palvelutarpeen omaavat henkilöt. Jos valikoitumista ei kontrolloida riittävästi, tutkimuksen tulokset voivat osoittaa, että palvelujen käyttö on yhteydessä heikentyneeseen terveyteen ja alentuneeseen hyvinvointiin. Tämänkaltaisten tulokset johtavat helposti päätelmään vaikuttamattomista palveluista, vaikka

havaittu käänteinen yhteys palvelujen käytön ja tulosmuuttujien välillä selittyikin koe- ja vertailuryhmän välisillä tarve-eroilla.

Seuraavassa esimerkissä 4 kuvataan valikoitumisharhan vaikutuksia ja kontrolloimista puolikokeellisella asetelmalla tehdyn nuorten osa-aikatyö -tutkimusesimerkin kautta (Monahan ym. 2011). Tämä esimerkki havainnollistaa myös sitä, kuinka koe- ja vertailuryhmät voidaan muodostaa silloin, kun satunnaistaminen ei ole mahdollista: koe- ja vertailuryhmän muodostamisessa hyödynnettiin tietoa siitä, oliko nuoren työmarkkinastatus muuttunut tutkimusaikana.

Esimerkki 4. Nuorten osa-aikatyö tutkimus

Nuorten osa-aikatyö -tutkimuksessa (Monahan, Lee & Steinberg 2011) toistettiin aiemmin julkaistun tutkimuksen (Steinberg, Fegley & Dornbusch 1993) analyysit käyttämällä valikoitumisharhan paremmin kontrolloivia analyysimenetelmiä. Tarkoituksena oli tutkia työunneissa tapahtuvien muutosten vaikutusta nuoren tilanteeseen (koulutus, käyttäytyminen ja psykososiaalinen tilanne). Tutkimuksen kohderyhmänä olivat USA:ssa, Wisconsinissa ja Kalifornian pohjoisosassa asuneet lukion ensimmäisen ja toisen vuosikursin opiskelijat (N=1 792). Aineisto kerättiin vuosina 1987–1988 ja 1988–1989 yhdeksästä koulusta, jotka valittiin erilaisilta alueilta.

Alkuperäisen analyysin mukaan koulun ohella tapahtuva intensiivinen työskentely (yli 20 h/vko) ennusti heikompaa akateemista menestystä vuoden seurantajaksolla. Kyseessä oli siis pitkittäisaineisto, jossa tutkijat kontrolloivat (*time-lagged controls* -menetelmällä) demografisia tekijöitä ja alkumittausten arvoja.

Myöhemmin havaittiin, että tulokset saattaisivat osin selittyä valikoitumisharhalla. Sen vuoksi toistetussa tutkimuksessa nuoret jaettiin koe- ja vertailuryhmään sen mukaan, oliko nuoren työskentelystatuksessa tapahtunut muutosta vai ei käyttäen *propensity score* -kaltaistamismenetelmää. Valikoitumisharhan kontrolloimisessa hyödynnettiin kaikkiaan 22 muuttujaa, jotka mitattiin nuorten aloittaessa yhdenntoista opiskeluvuotensa (aika 1; ks. tarkemmin Monahan, Lee & Steinberg 2011). Muuttujat liittyivät siihen, kuinka todennäköisesti nuori on työllinen tai työtön. Lisäksi tutkimusta toistettaessa puuttuvat tiedot korvattiin hyödyntäen moni-imputointia (*multiple imputation*).

Toistetussa tutkimuksessa havaittiin, että jos nuori ryhtyy tekemään enemmän kuin 20 tuntia osa-aikatyötä viikossa, hänen koulumenestyksensä heikkenee ja ongelmakäyttäytyminen lisääntyy. Kaksi eroavaisuutta suhteessa alkuperäiseen tutkimukseen tuli kuitenkin esille: enemmän työskennelleet nuoret eivät toistoanalyysin mukaan luottaneet heikommin itseensä. Lisäksi koko vuoden yli 20 tuntia viikossa työskennelleillä havaittiin heikohko keskiarvon kohoamistrendi, jota saattoi selittää se, että nämä opiskelijat hakeutuivat helpommille kursseille.

Lähde: Monahan ym. (2011)

Esimerkki 4 osoittaa siis sen, että valikoitumisharhan kontrolloimiseen ei havainnoivassa tutkimusasetelmassa riitä pelkästään tulosmuuttujiin (nuorten koulumenestys) vaikuttavien tekijöiden vakioiminen. Koska puolikokeellisesta tutkimuksesta puuttuu tutkittavien satunnaistaminen koe- ja vertailuryhmiin, tulee lisäksi vakioida mahdollisimman tarkasti ne tekijät, joiden vaikutuksesta tutkittavat valikoituvat joko koe- tai vertailuryhmään. Tässä tutkimuksessa päätulos pysyi samana valikoitumisharhan kontrolloimisen jälkeen, mutta joissain tapauksissa tulokset voivat vinoutua pahemminkin, jollei valikoitumisharhaa korjata analyysiteknisesti.

Valikoitumisharhan lisäksi muita tavallisimpia harhan lähteitä havainnoivissa tutkimusasetelmissa ovat **ajalliset muutokset** (*secular trends*), **väliin tulevat tapahtumat** (*interfering events*) ja **luonnollinen kehittyminen** (*maturation*) (Rossi ym. 2004, 272¹⁶). Näistä ajalliset muutokset viittaavat tilanteeseen, jossa yhteiskunnassa, alueella tai yhteisössä tapahtuvat muutokset edistävät tai heikentävät tutkitun palvelun todellista vaikutusta. Tällainen tilanne syntyy esimerkiksi silloin, kun tutkitaan työllisyystoimenpiteen vaikutusta työllisyyteen noususuhdanteessa, jolloin työllisyystrendi on kohentuva ilman erillisiä työllisyystoimiakin. Väliin tulevat tapahtumat puolestaan viittaavat tilanteeseen, jossa jokin tutkitusta palvelusta riippumaton tapahtuma vaikuttaa tutkimustuloksiin. Esimerkiksi jos tutkittaisiin hoitajamitoituksen vaikutusta vanhuspalvelujen laatuun, täytyisi huomioida myös tutkimusaikana terveydenhuollon organisaatiossa tapahtuvat muutokset ja niiden mahdolliset vaikutukset hoidon laatuun. Luonnollinen kehittyminen viittaa jo aiemmin sivuttuun ilmiöön eli siihen, kuinka osa muutoksista ei välttämättä johdu interventioista vaan luonnollisesti tapahtuvasta kehitymisestä. Esimerkiksi lasten lukutaitoa arvioitaessa kehitystä tapahtuu sekä luonnollisesti että tutkittavan intervention seurauksena, jolloin vaikutuksen

¹⁶ Katso lisätietoa sisäisen validiteetin uhkista esim. Robson (2001, 92–93) ”Käytännön arvioinnin perusteet” teoksesta

arvioiminen puhtaasti intervention tulokseksi voi tuottaa harhaisen estimaatin. (esim. Robson 2001, 92–93; Rossi ym. 2004, 272–273.)

Epävarmuustekijöiden kontrolloiminen
tilastollisten analyysimenetelmien avulla

Esittelemme seuraavaksi erilaisia tilastollisia analyysimenetelmiä, joiden avulla on mahdollista vähentää ja kontrolloida erilaisten tilastollisten harhojen, erityisesti valikoitumisharhan, mahdollisuutta. Näitä ovat: koe- ja vertailuryhmien kaltaistamisessa käytetyt menetelmät, kuten *propensity score*-menetelmä, regressiomenetelmät, regressioepäjatkuvuusmenetelmä, instrumenttimuuttujien käyttäminen ja *difference in difference* -analyysimenetelmä (ks. esim. Pekkarinen 2006; Blundell & Dias 2009; Loi & Rodrigues 2012). Tutkija voi esittämienne lähteiden avulla perehtyä näihin menetelmiin tarkemmin ja valita omaan vaikuttavuustutkimukseensa soveltuvan analyysistrategian.

Koe- ja vertailuryhmien kaltaistaminen

Havainnoivissa tutkimusasetelmissä ei ole satunnaisesti muodostettua koe- ja kontrolliryhmää, jolloin verrattavat ryhmät muodostetaan usein jälkikäteen kerätyn aineiston avulla. Koeryhmä saadaan tällöin useimmiten suoraan aineistosta (esim. kodin ulkopuolelle sijoitetut lapset), mutta vertailuryhmän muodostamiseksi joudutaan käyttämään jotain **kaltaistamismenetelmää** (*matching*), jotta ryhmät saataisiin vertailukelpoisiksi (ks. Rossi ym. 2004, 285–288). Yksinkertaisimmillaan kaltaistaminen voi tarkoittaa esimerkiksi sitä, että vertailuryhmään poimitaan henkilöitä, jotka ovat sukupuoli- ja ikäjakauman tai ammattiaseman mukaan samanlaisia kuin koeryhmään kuuluvat henkilöt.

Hienostuneempaa kaltaistamismenetelmää sen sijaan edustaa **propensity score -menetelmä** (ks. Rosenbaum & Rubin 1983, 1984;

Rossi ym. 2004, 285–288): Siinä koeryhmään kuulumisen todennäköisyyttä mallinnetaan aluksi logistisen regressioanalyysin avulla. Tämän jälkeen mallin avulla etsitään verrokkit, jotka eivät ole olleet toimenpiteen kohteena, mutta joilla on lähes yhtä suuri todennäköisyys päätyä intervention kohteeksi kuin koeryhmään kuuluvilla henkilöillä. Kaltaistetulla aineistolla voidaan lopuksi laskea valikoitumisharhasta puhdistetut estimaatit palvelun tai hoidon vaikutuksille. Menetelmän toimivuus valikoitumisharhan kontrolloinnissa riippuu siitä, miten hyvin ja missä määrin valikoitumista aiheuttavia tekijöitä kyetään mittaamaan, havaitsemaan ja mallintamaan. Jos merkittävä osa valikoitumisesta aiheutuu tutkijalle havaitsemattomista tekijöistä, menetelmän sovellettavuus valikoitumisharhan kontrolloimisessa heikkenee.

Empiirisissä tutkimuksissa *propensity score* -menetelmää on sovellettu esimerkiksi tutkittaessa lastensuojelun sijaishuollon asiakkuuden vaikutuksia koulutukseen, rikollisuuteen ja erilaisiin käyttäytymishäiriöihin (esim. Berger, Bruch, Johnson, James & Rubin 2009; Väisänen, Linnosmaa, Kestilä, Paananen & Gissler 2014) sekä nuorten työmarkkina-aseman vaikutuksia koulumenestykseen (ks. esimerkki 4, Monahan ym. 2011). Väisänen ja tutkijakumppanit (2014) esimerkiksi lähtevät liikkeelle havainnosta, jonka mukaan sijaishuollon entiset asiakkaat ovat matalammin koulutettuja ja he tekevät enemmän rikoksia kuin ikätoverinsa. Rikollisuus ja matala koulutus eivät kuitenkaan ole välttämättä seurausta sijaishuollon laadusta vaan asiakkaiden valikoitumisesta. Sen vuoksi sijaishuollon vaikuttavuuden arvioimiseksi he muodostivat sijaishuoltoon sijoitetuille vertailuryhmän ei-sijoitettujen joukosta. Tämä tapahtui lasquemalla todennäköisyydet lapsen kodin ulkopuoliselle sijoitukselle taustamuuttujien suhteen, minkä jälkeen ei-sijoitettujen joukosta haettiin vastinparit, joiden todennäköisyys joutua sijoitetuksi oli taustamuuttujien osalta lähes yhtä suuri kuin sijoitetuilla. Tutkimuksen tulokset kuitenkin osoittivat, että sijaishuolto ei keskimäärin ollut vaikuttavaa. Kodin ulkopuolisessa sijoituksessa olleiden nuorten

koulutus jäi matalammaksi ja he tekivät enemmän rikoksia kuin ikätoverinsa. (Väisänen ym. 2014.)

Regressiomalliin perustuvat menetelmät

Usein vastinpareja on kuitenkin hankala löytää, koska kaikkia olennaisia taustamuuttujia ei ole mitattu tai vastinpareja ei kyetä muodostamaan havaittujen muuttujien avulla. Jälkimmäisessä tilanteessa voidaan kuitenkin hyödyntää erilaisia regressiomallinnuksen tekniikoita. Regressiomalleihin voidaan ottaa joko aineistossa jo olevia, havaittuja, tai rekisteriaineistoista saatavia kontrollimuuttujia, joiden suhteen koe- ja vertailuryhmät eroavat, vaikka niiden perusteella ei muodostettaisikaan vastinpareja. Kontrollimuuttujien lisäämisellä pyritään vähentämään **puuttuvien muuttujien harhaa** (*omitted variable bias*). **Kontrollimuuttujien hyödyntäminen** ryhmien kaltaistamisessa onkin osin syrjäyttänyt vastinparien etsimisen (Rossi ym. 2004, 279), sillä kaltaistamiseen perustuvat menetelmät tyypillisesti vähentävät analyysissä käytettävien havaintojen lukumäärää. Joissakin tapauksissa, etenkin jos aineisto on suuri, koe- ja vertailuryhmien kaltaistaminen voi myös olla työlästä.

Kontrollimuuttujien käyttäminen on tarpeellista usein esimerkiksi palkkaeroja vertailtaessa. Peter Rossi kollegoineen (2004, 279–281) tarjoaa tästä kuvitteellisen esimerkin, jossa 35–40-vuotiaat miehet ovat osallistuneet työllisyysprojektiin. Ennen kontrollimuuttujien käyttämistä tulokset osoittavat, että projektiin osallistuneet tienasivat 0,45 dollaria vähemmän kuin ne, jotka eivät siihen osallistuneet. Rossi kollegoineen (2004, 281) kuitenkin toteaa, että tulokset ovat harhaisia, koska projektiin osallistuneiden joukossa on enemmän matalan koulutuksen saaneita. Siten koulutuksen kontrolloiminen tämänkaltaisissa tutkimuksissa on tarpeen. Käytännössä tutkimuksissa hyödynnetäänkin lähes aina kontrollimuuttujia, joiden tiedetään vaikuttavan tulosmuuttujaan eli tässä tapauksessa palkkoihin. Tästä huolimatta kontrollimuuttujien käyttöökään ei takaa sitä, että

valikoitumisharhan vaikutus tuloksiin poistuisi täysin. Esimerkiksi palkkaerojen tapauksessa on mahdollista, että projektiin osallistuneet ovat olleet motivoituneempia löytämään työtä projektin jälkeen, mikä voi selittää ryhmien välisiä palkkaeroja. Osallistujien motivaatiotasoa on kuitenkin hankala saada mallinnettua kontrollimuuttujien avulla. (Rossi ym. 2004, 281.)

Regressioepäjatkuvuus

Valikoitumisharhaa voidaan myös kontrolloida **regressioepäjatkuvuusmenetelmän** (*regression discontinuity design*) avulla, jota voidaan käyttää, jos tutkittuun palveluun osallistuminen riippuu ratkaisevasti tietyin jatkuvan muuttujan kynnyksarvosta (ks. esim. Thistlethwaite & Campbell 1960; Venkataramani, Bor & Jena 2016). Tällöin voidaan olettaa, että tutkittavat ihmiset juuri kynnyksarvon alapuolella ovat havaittavilta ja havaitsemattomilta ominaisuuksiltaan samanlaisia kuin juuri kynnyksarvon yläpuolelle jäävät ihmiset. Siten arvioimalla tulosmuuttujien arvoja kynnyksarvon ala- ja yläpuolelle jäävien tutkittavien osalta voidaan arvioida intervention vaikuttavuutta.

Regressioepäjatkuvuusmenetelmää on sovellettu esimerkiksi tutkittaessa, onko lihavuudella vaikutusta palkkatuloihin (Levasseur 2019). Pierre Levasseur (2019) käytti painoindeksille määritellyjä viitearvoja jakaakseen tutkittavat ylipainon ($BMI=25 \text{ kg/m}^2$) ja liikalihavuuden ($BMI=30 \text{ kg/m}^2$) kynnyksarvojen ylä- ja alapuolelle jääviin, jonka jälkeen hän vertaili ryhmien välisiä palkkaeroja. Keskeisenä oletuksena tutkimuksessa oli, että BMI:n raja-arvot toimivat satunnaistavina tekijöinä, jolloin raja-arvojen ympäristöön sijoittuvat henkilöt ovat samanlaisia taustamuuttujien osalta muodostaen luonnollisen koe- ja vertailuryhmäasetelman. Tätä taustaa vasten tutkimuksessa voitiinkin päätellä, ettei ylipaino ($BMI>25 \text{ kg/m}^2$) itsessään toimi Meksikon työmarkkinoilla palkkatuloon vaikuttavana tekijänä, mutta siirtyä ylipainosta liikalihavuuden ($BMI>30 \text{ kg/m}^2$) puolelle vähentää tuntipalkkaa noin 15 prosentilla.

Instrumenttimuuttujamenetelmä

Regressioepäjatkuvuusmenetelmän käyttö ei ole kuitenkaan aina mahdollista, sillä kaikissa tutkimuksissa ei voida selkeästi määritellä jatkuvaan muuttujaan perustuvaa kynnyksarvoa. Tällöin voidaan turvautua muihin menetelmiin, kuten **instrumenttimuuttuja (IV) -menetelmään**. Tätä menetelmää voidaan käyttää, jos valikoituminen koe- tai vertailuryhmään riippuu kolmannesta, mallin ulkopuolisesta instrumenttimuuttujasta. Instrumenttimuuttujan täytyy korreloida interventioon osallistumisen kanssa ilman, että se vaikuttaisi suoraan tulosmuuttujaan. Toisin sanoen instrumenttimuuttuja vaikuttaa tutkittavaan tulosmuuttujaan ainoastaan interventioon osallistumisen välityksellä. Tällainen tilanne tulee usein vastaan niin sanotuissa ”luonnollisissa koetilanteissa”, joissa palveluun osallistuminen voi muistuttaa satunnaista johtuen IV-menetelmän soveltamisesta. (Pekkarinen 2006.)

Esimerkiksi Joseph Doyle (2013) kuvaa tutkimuksessaan hyvin samankaltaista tilannetta kuin aiempi *propensity score* -menetelmää koskeva esimerkkimme (Väisänen ym. 2014). Molemmissa tutkimuksissa tarkastelun kohteena on kodin ulkopuolelle sijoittamisen vaikutus riskiin, että lapset päätyisivät nuorisorikollisiksi. Doyle (2013) nostaa esille, että *propensity score* -menetelmä ei välttämättä ole riittävä tutkittaessa sijaishuollon vaikuttavuutta. Menetelmässä oletetaan, että todennäköisyys tulla sijoitetuksi kodin ulkopuolelle ei riipu havaitsemattomista tekijöistä, mikä ei välttämättä pidä paikkaansa käytännön sovelluksissa. Hän esittää sen sijaan, että sijoituksen todennäköisyyteen vaikuttaa se, sattuuiko lapselle ”tiukka” sosiaalityöntekijä, joka tekee sijoituspäätöksiä helpommin vai ”lempeä” sosiaalityöntekijä, joka ei vastaavassa tilanteessa olisi sijoittanut lasta kodin ulkopuolelle. Tällöin sosiaalityöntekijän valikoituminen toimisi satunnaistavana tekijänä tuottaen samalla niin sanotun luonnollisen koetilanteen. Tilanne täyttää instrumenttimuuttujalle asetetut ehdot: sosiaalityöntekijän valikoituminen vaikuttaa sijoitus-

päätökseen, mutta sillä ei ole suoraa yhteyttä tutkittavaan tulosmuuttajaan. Instrumenttimuuttujan käytöstä huolimatta Doylen (2013) tulokset sijoitusten rikollisuutta ehkäisevästä vaikutuksesta eivät olleet kuitenkaan imartelevia, sillä kodin ulkopuolelle sijoitetuilla lapsilla oli edelleen suurempi todennäköisyys päätyä nuorisorikollisiksi kuin niillä lapsilla, joita ei ollut sijoitettu kodin ulkopuolelle.

Instrumenttimuuttujan käyttöön liittyy kuitenkin sellainen rajoitus, että se pätee ainoastaan tiettyihin marginaalitapauksiin, joihin instrumenttimuuttuja vaikuttaa. Esimerkiksi edellisessä esimerkissä instrumenttimuuttujalla kyetään arvioimaan nuorisorikollisuuden todennäköisyyttä niille nuorille, joiden tilanne sijoittamisajankohdaksi oli sellainen, että sijoituspäätös voi riippua sosiaalityöntekijän taipumuksesta tehdä kodin ulkopuolisia sijoituspäätöksiä. Sen sijaan malli ei kykene estimoimaan nuorisorikollisuuden todennäköisyyttä niille nuorille, jotka olisi sijoitushetkenä sijoitettu tai jätetty sijoittamatta kodin ulkopuolelle joka tapauksessa, riippumatta sosiaalityöntekijän taipumuksesta tehdä sijoituspäätöksiä. (Doyle 2013.)

Difference-in-differences -menetelmä

Lopuksi esitellään vielä **difference in difference (DID) eli erotuksen erotus -analysointimenetelmä**, joka on erittäin tehokas, kun valikoitumista koe- ja vertailuryhmiin aiheuttaa ajasta riippumaton havaitsematon tekijä. Menetelmän käytön edellytyksenä on, että tutkittavia on mahdollista seurata tietyn ajan verran. Menetelmää käytetään siten, että sekä koe- että vertailuryhmiä seurataan ennen interventiota (t_0) ja sitä seuraavalla ajanjaksolla, jolloin koeryhmä on käyttänyt tutkimuksen kohteena ollutta palvelua (ja vertailuryhmä ei ole saanut tätä palvelua) (t_1). Tällöin vähentämällä t_1 ajankohdassa mitatuista tulosmuuttujan arvoista t_0 ajankohdan arvot saadaan molemmista ryhmistä poistettua kiinteiden ja ajassa-muuttumattomien havaitsemattomien tekijöiden vaikutus tulosmuuttujaan. Tämän jälkeen koe- ja vertailuryhmien erotuksen vertailu taas poistaa ryhmille

yhteisen aikatrendin vaikutuksen intervention lopputulokseen. (ks. esim. Pekkarinen 2006; Loi & Rodrigues 2012, 32–38; Wing, Simon & Bello-Gomez 2018.)

Erotuksen erotus -analysointimenetelmää on hyödynnetty esimerkiksi saksalaisessa Michael Gebelin ja Jonas Voßemerin (2014) tutkimuksessa, jossa tarkasteltiin kuinka siirtymät työllisyyden ja työttömyyden välillä vaikuttavat terveyteen. Tutkimuksessaan he hyödynsivät *propensity score* -menetelmää ja erotuksen erotus -menetelmää (ns. DID-PMS: ks. Heckman, Ichimura & Todd 1997), jolloin he oletettavasti kykenivät kontrolloimaan niin havaittuja kuin havaitsemattomia terveyteen vaikuttavia tekijöitä. Tutkijat kaltaistivat koe- ja vertailuryhmät *propensity score* -menetelmän avulla. Siten voitiin olettaa, että erotuksen erotus -menetelmään olennaisesti kuuluva oletus yhdenmukaisista ajallisista trendeistä koe- ja vertailuryhmän välillä (*parallel trends*) pitää paikkansa. Tällöin esimerkiksi ikääntymisen vaikutus koe- ja vertailuryhmän terveyteen on samanlainen. Tämän jälkeen Gebel ja Voßemer (2014) poistivat erotuksen erotus -menetelmää hyödyntämällä kiinteät havaitsemattomat vaikutukset koe- ja vertailuryhmistä ja ajasta (kuten ikääntymisestä) johtuvat vaikutukset terveydentilassa havaittavaan trendiin. Tulosten perusteella Gebel ja Voßemer (2014) kykenivät osoittamaan, että muutokset työllisyyden ja työttömyyden välillä vaikuttavat yksilön psyykkiseen, mutta eivät fyysiseen terveyteen. Fyysinen terveys kuitenkin vaikuttaa heidän analyysinsa mukaan työllistymisen todennäköisyyteen, minkä vuoksi myös fyysisen terveyden edistäminen olisi tärkeää aktivointipoliittisesta näkökulmasta ajateltuna.

Olemme edellä esitelleet erilaisia tilastollisia analyysitekniikoita, joita käyttämällä hoitojen ja palvelujen vaikuttavuutta voidaan tutkia erilaisilla havainnoivillakin tutkimusasetelmilla mahdollisimman luotettavasti. Jos erilaiset harhatekijät saadaan onnistuneesti kontrolloitua esimerkiksi edellä esiteltyjen tilastollisten menetelmien avulla, voidaan siis myös korkealaatuisilla havainnoivilla tutkimuksilla tuottaa luotettavaa ja käyttökelpoista tietoa, jonka avulla voidaan

kehittää sosiaali- ja terveyspalveluja käytännössä ja palvelujärjestelmätasolla. Vaikuttavuustutkimuksen uskottavuuden näkökulmasta on tärkeää, että tutkija raportoi huolellisesti sen, oliko koe- ja vertailuryhmien välillä alkumittauksessa tilastollisesti merkitsevää eroa, ja jos oli, saatiinko ryhmät toisilleen verrannollisiksi analyysitekniisin keinoin. Tutkimustuloksia raportoitaessa on myös kerrottava ne käytetyt tutkimusmenetelmät ja kohennustoimenpiteet, joiden avulla pyrittiin vähentämään tutkimustuloksiin vaikuttavia harhatekijöitä. On kuitenkin hyvä huomioida se, että havainnoivissa tutkimusasetelmissä ei koskaan voida täysin poissulkea sitä riskiä, että jokin muu tekijä, jota tutkimuksessa ei ole pystytty mittaamaan ja kontrolloimaan, voisi ollakin havaittua yhteyttä selittävä, niin kutsuttu **sekoitava tekijä** (*confounding factor*) (Rubin & Babbie 2008, 256). Tämä ”vastuuvapauslauseke” on tutkijan hyvä joka tapauksessa siis mainita havainnoivan tutkimuksensa tutkimusraportissa.

Kustannusvaikuttavuuden määrittely ja tutkiminen

Kustannusten määrittely ja mittaaminen

Toimenpiteiden tai palvelujen kustannusvaikuttavuuden arvioimiseksi tarvitaan tietoa terveys- ja hyvinvointivaikutusten ohella myös niiden kustannuksista: mitä kustannuksia arvioinnissa huomioidaan ja miten näitä kustannuksia mitataan. Tuotantokustannusten mittaamiseen tarvitaan tietoa sekä tuotoksen (kuviokuva 1) tuottamiseen tarvittavista panoksista (tai resursseista), kuten työvoimasta, koneista, laitteista tai materiaaleista, sekä näiden hinnoista. Taloustieteellisessä tutkimuksessa panosten ja tuotoksen määrien välisten suhteiden tarkastelussa usein käytetään niin sanottua **tuotantofunktiota**, jonka avulla voidaan arvioida tuotoksen tuottamiseen tarvittavien panosten määriä. Panosten hintana puolestaan suositellaan käytettäväksi **vaihtoehtokustannusta**, joka mittaa resurssin arvoa seuraavaksi parhaassa vaihtoehtoisessa käytössä. Esimerkiksi omaishoitajien

hoivan tarjoamiseen käytetyn ajan vaihtoehtokustannusta voidaan mitata palkalla, jonka omaishoitajat ansaitisivat, jos he osallistuisivat palkkatyöhön omaishoidon sijasta (Van den Berg, Bleichrodt & Eeckhoudt 2005). Käytännössä kustannusten laskemisessa käytetyt hinnat eivät useinkaan vastaa vaihtoehtokustannusta, mutta hintoja voidaan yrittää korjata, että ne vastaisivat paremmin resurssien käytön taloudellista arvoa (esim. Brouwer, Rutten & Koopmanscharp 2010).

Palvelujen tuotannosta aiheutuvia kustannuksia voidaan luokitella sekä nykyisiin että tuleviin palvelujärjestelmän kustannuksiin ja järjestelmän ulkopuolisiin kustannuksiin (Brouwer ym. 2010). Palvelujärjestelmän kustannuksilla tarkoitetaan sosiaali- ja terveyspalvelujen ja mahdollisesti muiden olennaisten palvelujen tuotannosta syntyviä nykyisiä tai tulevia kustannuksia. Palvelujärjestelmän ulkopuolisiin kustannuksiin lasketaan potilaiden ja asiakkaiden palvelujen käyttämiseen tarvitseman ajan arvo ja tuottavuuskustannukset, joilla tarkoitetaan sairastumisesta aiheutuvia tuotannonmenetyksiä tai menetetyn vapaa-ajan arvoa (Schulpher 2010). Palvelujärjestelmän ulkopuolisia kustannuksia ovat myös esimerkiksi omaishoitajien ja läheisauttajien hoivan tarjonnasta syntyvät aikakustannukset.

Käytännön kysymys taloudellisessa arvioinnissa on, mitä **kustannuksia** palvelujen kustannusvaikuttavuuden arvioinnissa huomioidaan. Taloustieteessä suositellaan huomioimaan kaikki palvelujen tuotannosta aiheutuvat kustannukset riippumatta siitä, kenelle tai milloin ne aiheutuvat. Tätä näkökulmaa kustannusten laskentaan kutsutaan **yhteiskunnalliseksi näkökulmaksi**. Koska yhteiskunnallinen näkökulma voi osoittautua haasteelliseksi, usein käytännön sovelluksissa käytetään **palvelujärjestelmän näkökulmaa** (esim. Edwards, Cézillechair, Bywater, Hughes & Hutchings 2007), joka huomioi ainoastaan sosiaali- ja terveydenhuollon tai muiden olennaisten sektoreiden palvelujen tuottamisesta aiheutuvat kustannukset.

Kustannusvaikuttavuuden määrittely ja arviointi

Palvelujen kustannusvaikuttavuuden arvioinnissa tuotetaan tietoa palvelujen vaikuttavuudesta ja kustannuksista päätöksenteon tueksi. Kustannusvaikuttavuustutkimus edellyttää siis tietoa palvelujen vaikuttavuudesta, ja kustannusvaikuttavuusanalyysi voidaan myös yhdistää vaikuttavuustutkimukseen (ks. esim. O'Neill ym. 2013). Kustannusvaikuttavuustutkimuksen lähtökohta on tuottaa tietoa päätöksentekijöille, jotka pyrkivät tuottamaan mahdollisimman paljon terveyttä tai hyvinvointia käytössään olevilla rajallisilla voimavaroilla.

Kustannusvaikuttavuus on suhteellinen käsite, ja sen arviointi vaatii tietoa vähintään kahden palvelun tai hoidon kustannuksista ja vaikuttavuudesta. Kustannusvaikuttavuuden arvioinnissa erotellaan toisistaan tilanteet, joissa päätöksentekijä kohdentaa rajallisia voimavaroja joko keskenään kilpaileviin (tai vaihtoehtoiisiin) tai ei-kilpaileviin (tai riippumattomiin) hoitoihin tai palveluihin (esim. Meltzer & Smith 2011). Esimerkkeinä keskenään kilpailevista hoidoista voisivat olla simvastatiinin tai rosuvastatiinin käyttö kolesterolin hoidossa. Ei-kilpailevia hoitoja/palveluja voisivat olla verenpaineen hoidossa käytetyt terveystalvet ja ikääntyneelle henkilölle tarjotut arkipäivän selviytymistä tukevat sosiaalipalvelut.

Tarkastellaan seuraavaksi kustannusvaikuttavuuden arviointia tilanteessa, jossa terveydenhuollon päätöksentekijä kohdentaa rajallisia voimavaroja kahden vaihtoehtoisen hoidon, nykyisen ja uuden hoidon, tuottamiseen. Olkoon hoidon t terveyshyödyt (tai hyvinvointihyödyt sosiaalihuollossa) ja kustannukset E_t ja C_t , missä $t = 0$ viittaa nykyiseen hoitoon ja $t = 1$ uuteen hoitoon. Oletetaan, että päätöksentekijä tuntee hoitojen kustannukset ja terveyshyödyt. Tältä pohjalta hän arvioi hoitojen kustannusvaikuttavuutta ja päättää, kumpi hoito otetaan käyttöön.

Taulukossa 1 luokitellaan tilanteita, joihin sosiaali- ja terveystalvelujen päätöksentekijä voi joutua. Jos uusi hoito on samanhintainen

*Sosiaali- ja terveyspalvelujen vaikuttavuuden ja
kustannusvaikuttavuuden tutkiminen*

tai halvempi (sarakeet 2 ja 3 taulukossa 1) ja vähintään yhtä vaikuttava kuin vanha hoito (rivit 1 ja 2), olisi päätöksentekijän järkevää käyttää voimavarat uuteen hoitoon. Jos taas uusi hoito maksaa saman verran tai enemmän kuin vanha hoito (sarakeet 1 ja 2) ja sen tuottamat terveyshyödyt ovat samat tai pienemmät kuin vanhan hoidon (rivit 2 ja 3), päätöksentekijän kannattaisi pitäytyä vanhan hoidon tuottamisessa.

Taulukko 1. Kustannusvaikuttavuuden arviointi

	Sarake 1: $C_1 > C_0$	Sarake 2: $C_1 = C_0$	Sarake 3: $C_1 < C_0$
Rivi 1: $E_1 > E_0$	A	1	1
Rivi 2: $E_1 = E_0$	0	1 tai 0	1
Rivi 3: $E_1 < E_0$	0	0	B

Kustannusvaikuttavuuden tulkinnessa on kysymys eri vaihtoehtojen välisestä arvottamisesta, jonka perusteena on olennaisesti **päätöksentekijän maksuhalukkuus**. Päätös esimerkiksi uuden hoidon valitsemisesta on kaikista kiistattomin silloin, kun uusi hoito on sekä edullisempi että vaikuttavampi kuin vanha hoito. Solut A ja B (taulukko 1) ovat päätöksentekijän kannalta haastavampia. Jos uusi hoito on vaikuttavampi ja kalliimpi kuin vanha hoito (solu A), tulee päätöksentekijän arvioida omaa maksuhalukkuuttaan tuotetusta lisäterveydestä. Koska hoitojen kustannukset ja terveyshyödyt ovat tiedossa, voidaan päätöksenteon avuksi laskea **inkrementaalinen kustannusvaikuttavuussuhde** (*incremental cost-effectiveness ratio, ICER*):

$$ICER = \frac{C_1 - C_0}{E_1 - E_0},$$

Inkrementaalinen kustannusvaikuttavuussuhde mittaa tuotetun lisäterveyden (tai lisähyvinvoinnin) hintaa siirryttäessä vanhasta hoidosta uuteen hoitoon (Drummond, Schulpher, Claxton, Stoddard & Torrance 2015). Jos päätöksentekijän maksuhalukkuus tuotetusta terveydestä on vähintään yhtä suuri kuin inkrementaalinen kustannus-vaikuttavuussuhde, voidaan uutta hoitoa pitää kustannusvaikuttavana, ja päätöksentekijän on järkevää käyttää niukat voimavarat uuteen hoitoon vanhan hoidon sijasta.

Jos uuden hoidon tuottamat terveyshyödyt ovat pienemmät kuin vanhan hoidon terveyshyödyt ja uusi hoito on halvempi kuin vanha hoito (ks. taulukko 1, solu B), uuteen hoitoon siirtyminen säästää, mutta tuottaa vähemmän terveyshyötyä kuin vanha hoito. Vaikka uusi hoito ei ole niin vaikuttava kuin vanha, päätöksentekijä voi olla halukas ottamaan sen käyttöön, jos sen aiheuttamat kustannussäätöt ovat riittävän suuret suhteessa menetettyihin terveyshyötyihin. Tämä tarkoittaa tietenkin sitä, että päätöksentekijä samalla hyväksyy sen, että uusi hoito tuottaa vähemmän terveyttä ja hyvinvointia kuin vanha hoito.

Havainnollistamme hoidon kustannusvaikuttavuuden tutkimista ja siihen perustuvaa päätöksentekoa esimerkillä 5, joka perustuu kokonaisvaltaisen geriatrisen arvioinnin (CGA) kustannusvaikuttavuusanalyysiin (Lundqvist, Alwin, Henriksson, Husberg, Carlsson & Ekdahl 2018). Koska CGA sekä maksaa enemmän että on vaikuttavampi kuin tavanomainen hoito, on päätöksentekijän arvioitava maksuhalukkuuttaan CGA:lla tuotetusta lisäterveydestä. Jos inkrementaalisen kustannusvaikuttavuuden kaavaa sovelletaan esimerkin 5 aineistoon (taulukko E5.1), voidaan todeta, että CGA:n inkrementaalinen kustannusvaikuttavuussuhde on hieman yli 46 000 euroa (ks. myös Lundqvist ym. 2018). Uusi hoito on kustannusvaikuttava, jos päätöksentekijän maksuhalukkuus ylittää CGA:n inkrementaalisen kustannusvaikuttavuussuhteen.

*Sosiaali- ja terveyspalvelujen vaikuttavuuden ja
kustannusvaikuttavuuden tutkiminen*

Esimerkki 5 Kokonaisvaltaisen geriatrisen arvioinnin kustannusvaikuttavuuden tutkiminen

Lundqvist ja tutkijaryhmä (2018) tutkivat kokonaisvaltaisen geriatrisen arvioinnin (CGA) kustannusvaikuttavuutta suhteessa tavanomaiseen hoitoon. Tutkimuksen kohdeväestöön kuuluivat monisairaajat yli 75-vuotiaat Ruotsin Norrköpingissä asuvat vanhukset, joilla oli vähintään kolme sairaalan vuodeosastohoitojaksoa viimeisen 12 kuukauden aikana. Kokonaisvaltaiseen geriatriseen arviointiin kuului hoitajan ja sosiaalityöntekijän kotikäynti, jonka aikana arvioitiin asiakkaan ravitsemusta ja sosiaalista ympäristöä. Lisäksi farmaseutti, toimintaterapeutti ja fysioterapeutti arvioivat asiakkaiden lääkityksen ja toimintakyvyn. Koeryhmään kuuluville potilaille tehtiin henkilökohtainen palvelusuunnitelma geriatrian poliklinikalla. Kontrolliryhmään kuuluvat iäkkäät henkilöt saivat tavanomaista hoitoa.

Kokonaisvaltaisen geriatrisen arvioinnin pitkän aikavälin vaikutuksia palvelujen kustannuksiin ja elämänlaatuun arvioitiin Markov-mallin avulla. Saatujen tulosten mukaan kokonaisvaltaisen geriatrisen arviointi maksaa enemmän, mutta tuottaa enemmän laaturapainotettuja elinvuosia kuin tavanomainen hoito (taulukko E5.1).

Taulukko E5.1 Hoitojen kustannukset ja elämänlaatu, mukailten Lundqvist ym. 2018, taulukko 2

	CGA	Tavanomainen hoito
Kustannukset:		
Perusterveydenhuolto	7 881€	8 477€
Polikliininen hoito ^a	24 569€	9 696€
Vuodeosastohoito	21 854€	21 381€
Kunnalliset palvelut	59 024€	49 095€
Kustannukset yhteensä	113 327€	88 649€
Elämänaatu:		
QALY:t	3,33	2,80

^a geriatrisen ja muu polikliininen hoito yhdistetty

Lähde: Lundqvist ym. (2018).

Joissakin maissa sovelletaan kansallisia kustannusvaikuttavuuden kynnysarvoja uusien hoitojen tai menetelmien kustannusvaikuttavuuden arvioinnissa. Esimerkiksi Englannissa suositellaan käytettäväksi terveydenhuollon hoitoja, joiden inkrementaalinen kustannus-vaikuttavuussuhde on enintään 30 000 puntaa¹⁷ per niiden tuottama QALY (Drummond ym. 2015). Tämänkaltaiset kansalliset suositukset asettavat ylärajan tuotetun terveyshyödyn hinnalle ja ohjeistavat yksittäisiä terveydenhuollon päätöksentekijöitä heidän arvioissaan terveydenhuollon menetelmien kustannuksista ja hyödyistä.

Jos taloudellisen arvioinnin kohteena ovat ei-kilpailevat hoidot, joiden vaikuttavuutta voidaan mitata yhteismitallisesti, päätöksentekijä voi siis järjestää hoidot **kustannusvaikuttavuussuhteen** (kustannukset per terveyshyödyt) perusteella paremmuusjärjestykseen ja kohdentaa rajalliset voimavarat tehokkaimpien hoitojen tuottamiseen (esim. Meltzer & Smith 2011).

Vaikuttavuus- ja kustannusvaikuttavuustutkimusten tulosten hyödyntämisen mahdollisuudet uudistuvalla sosiaali- ja terveysalalla

Olemme tässä artikkelissa käsitelleet vaikuttavuuden ja kustannusvaikuttavuuden mittaamista, vaikuttavuustutkimukseen soveltuvia tutkimusasetelmia ja erilaisia analyysitekniisiä mahdollisuuksia erityisesti valikoitumisharhan kontrolloimiseen. Artikkelimme keskeinen viesti vaikuttavuustutkimusta ja kustannusvaikuttavuustutkimusta suunnitteleville on se, että tutkimukset on syytä suunnitella ja toteuttaa huolellisesti, jotta tutkimustulokset olisivat mahdollisimman luotettavia ja hyvin hyödynnettävissä päätöksenteossa.

Tutkijan on tärkeää valita luotettavat tulosmittarit sekä mahdollisimman vahva tutkimusasetelma. Vaihtoehtoisia tutkimusasetelmia ovat kokeelliset ja havainnoivaan aineistoon perustuvat asetelmat. Hoidon ja palvelujen vaikuttavuuden arvioinnissa on olennaista, että

¹⁷ 30 000,00 GBP = 34 473,84 EUR, kun 1 GBP = 1,149128 EUR (<https://valuuttamuunnin.com/> 28.4.2020)

tutkimuksissa pystytään tyydyttävästi vastaamaan vaikuttavuuden arvioinnin peruskysymykseen ja mittaamaan hoidon ja palvelujen vaikuttavuus mahdollisimman luotettavasti. Käytännön syiden takia usein päädytään käyttämään havainnoivia aineistoja ja havainnoivaa tutkimusasetelmaa täydennetään erilaisilla tilastollisilla analyysimenetelmillä, joilla kontrolloidaan erilaisia vaikuttavuuden arvioinnin harhalähteitä, erityisesti valikoitumisharhaa. Kun erilaiset harhojen lähteet onnistutaan huolellisen tutkimusprosessin ja analyysitekniesten keinojen avulla kontrolloimaan tyydyttävästi, myös havainnoivaan aineistoon perustuvalla tutkimusasetelmalla voidaan tuottaa palvelujärjestelmän ja käytäntöjen kehittämisen kannalta luotettavia tuloksia.

Vaikuttavuus- ja kustannusvaikuttavuustutkimusten erityisyys tieteellisen tutkimuksen kentässä on se, että niiden tavoitteena on juuri tukea sosiaali- ja terveyspalvelujen käytäntöä ja päätöksentekoa. Tutkimustulosten sovellettavuutta käytännössä kuitenkin vaikeuttaa tutkimuksen ja käytännön erilainen aikajänne. Sosiaali- ja terveysalan ammattilainen tai päättäjät tarvitsee työssään ja päätöksenteossa välittömästi tai mahdollisimman pikaisesti hyödynnettävää tietoa. Erityisesti laadukkaana kokeellisen tutkimuksen toteuttaminen on usein kuitenkin hidasta ja kallista. Tieteelliset tutkimustulokset yleensä julkaistaan vertaisarvioituissa lehdissä, joiden arviointiprosessit voivat kestää kauan. Tämä voi edelleen hidastaa tutkimustulosten välittämistä käytännön toimijoiden tietoon. Esimerkiksi näiden syiden takia ei ole itsestään selvää, että käytössä olevien sosiaali- ja terveydenhuollon palvelujen vaikuttavuus tai kustannusvaikuttavuus olisi todennettu tutkimuksen keinoin.

Yksi mahdollisuus tuottaa luotettavaa vaikuttavuus- ja kustannusvaikuttavuustietoa sosiaali- ja terveysalan kehittämisen tueksi on parantaa mahdollisuuksia puolikokeellisten tutkimusasetelmien käyttämiseen. Laadukkaat valmiit ja tutkimuskäyttöön helposti saatavilla olevat aineistot mahdollistavat puolikokeellisten takautuvien tutkimusasetelmien käyttämisen. Puolikokeellisen tutkimusasetel-

man toteuttaminen yhdessä asianmukaisten analyysimenetelmien kanssa tuottaa soveltamiskelpoista tietoa ja on nopeampaa kuin korkeellisten tutkimusten tekeminen. Tässä suhteessa Suomessa tilanne on jo suhteellisen hyvä, koska suomalaista palvelujärjestelmää tarkastelevassa tutkimuksessa hyödynnetään laajasti lakisääteisiä rekisteriaineistoja, kuten esimerkiksi Kelan lääkekorvausrekisteriä ja THL:n ylläpitämää hoitoilmoitusrekisteriä (Hilmo).

Vaikuttavuus- ja kustannusvaikuttavuustutkimuksen nykyistä ripeämpää toteuttamista edistäisivät tietenkin myös nykyistä laadukkaammat ja tietosisällöltään rikkaammat rekisteriaineistot, jotka soveltuisivat myös tutkimuskäyttöön. Olennainen edellytys tähän on erityisesti se, että rekisteritiedon keskitetty kirjaaminen toissijaisiin käyttötarkoituksiin etenee ja yleistyy valtakunnallisesti ja kattavasti sekä terveyden- että sosiaalihuollon raportointikäytänteisiin. Selkeä edistysaskel tässä suhteessa on Toisiolaki (L552/2019), joka mahdollistaa potilas- ja asiakasasiakirjoihin kirjattujen tietojen käytön asiakirjarakenteiden välityksellä myös erilaisiin toissijaisiin käyttötarkoituksiin, kuten tieteelliseen tutkimukseen, kehittämistoimintaan, opetukseen ja tietojohdamiseen. Sosiaali- ja terveydenhuollon tietolupaviranomainen Findata vastaa toisiolain tarpeisiin toimien eri tahoja yhdistävänä voimana ja tuottaen anonymisoitua rekisteritietoa tutkijoiden käyttöön. Terveydenhuollon kansallisiin laaturekistereihin on kaavailtu pohjaksi myös Kanta-järjestelmän kautta keskitettyä tiedonkeruuta, joka toimisi niin, että tarvittavat rekisteritiedot voitaisiin poimia jatkossa Kanta-tietoaltaasta soveltuvilla poimintaehdoilla (ks. Jonsson ym. 2019, 144–145). Erityisesti sosiaalihuollon tietoinfrastruktuuria ja kirjaamiskäytäntöjä on vielä kehitettävä (Väyrynen, Ailio, Arajärvi & Forsell 2020, 4–5).

On toivottavaa, että lisääntyviä mahdollisuuksia hoitojen ja palvelujen vaikuttavuustiedon tuottamiseen ryhdytään hyödyntämään enenevässä määrin sosiaali- ja terveydenhuollon toiminnassa. Tämä vaatii toiminnan aktiivista johtamista vaikuttavuustiedon lisääntyvän tuottamisen ja hyödyntämisen suuntaan.

Lähteet

- Berger Lawrence M., Bruch Sarah K., Johnson Elizabeth I., James Sigrud & Rubin David 2009. Estimating the “Impact” of Out-of-Home Placement on Child Well-Being: Approaching the Problem of Selection Bias. *Child Development* 80(6), 1856–1876.
- Blundell Monica & Dias Costa 2009. Alternative Approaches to Evaluation in Empirical Microeconomics. *Journal of Human Resources* 44(3), 565–640.
- Brooks Richard 1996. EuroQol: the current state of play. *Health Policy*, 37(1), 53–72.
- Brouwer Werner, Rutten Frans & Koopmanschap Marc 2010. Costing in economic evaluations. Teoksessa Drummond Michael & McGuire Alistair (toim.) *Economic evaluation in health care. Merging theory with practice*. Oxford University Press, Oxford, 68–93.
- Chen Huye-Tsyh 2005. *Practical program theory. Assessing and improving planning, implementation, and effectiveness*. Sage, United States of America, California, Thousand Oaks.
- Craig Peter, Dieppe Paul, Macintyre Sally, Michie Susan, Nazareth Irwin & Petticrew Mark 2008. *Developing and evaluating complex interventions*. Medical Research council. <https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/>
- Dahler-Larsen Peter 2005. *Vaikuttavuuden arviointi*. Hyvät käytännöt menetelmä-käsikirja. FinSoc arviointiraportteja 3. Stakes, Helsinki.
- Doyle Joseph 2013. Causal effects of foster care: An instrumental-variables approach. *Children and Youth Services Review* 35(7), 1143–1151.
- Drummond Michael, Schulpher Mark, Claxton Karl, Stoddard Greg & Torrance George 2015. *Methods for the Economic Evaluation of Health Care Programmes*. 4. painos. Oxford University Press, Oxford.
- Edwards Rhiannon, Cèilleachair Alan, Bywater Tracey, Hughes Dyfrig & Hutchings Judy 2007. Parenting programme for parents of children at risk of developing conduct disorder: cost-effectiveness analysis. *British Medical Journal* 334(7595), 682.
- Fraser Mark 2009. *Intervention research: developing social programs*. Chapter 1. What is intervention research? Oxford University Press, Oxford, New York.
- Gamerman Victoria, Cai Tianxi & Elsäßer Amelie 2018. Pragmatic randomized clinical trials: best practices and statistical guidance. *Health Services and Outcomes Research Methodology* 19(1), 23–35.
- Gebel Michael & Voßemerin Jonas 2014. The impact of employment transitions on health in Germany. A difference-in-differences propensity score matching approach. *Social Science & Medicine* (1982) 108, 128–136.

- Gottfredson Denise, Cook Thomas, Gardner Frances, Gorman-Smith Deborah, Howe George, Sandler Irwin & Zafft Katryn 2015. Standards of Evidence for Efficacy, Effectiveness, and Scale-up Research in Prevention Science: Next Generation. *Prevention Science* 16(7), 893–926.
- Heckman James, Ichimura Hidehiko & Todd Petra 1997. Matching as an econometric evaluation estimator. Evidence from evaluating a job training programme. *Review of Economic Studies* 64(4), 605–654.
- Ikonen Tuija & asiantuntijaryhmä 2019. Asiakasryhmäkohtainen tieto laadusta ja vaikuttavuudesta sosiaali- ja terveydenhuollon tiedolla johtamisessa ja ohjauksessa. Laaturekisterien asema palvelujärjestelmässä. Sosiaali- ja terveysministeriön raportteja ja muistioita 70, Helsinki. <http://urn.fi/URN:ISBN:978-952-00-4129-8>
- Jonsson Pia Maria, Pikkujämsä Sirkku & Heiliö Piia-Liisa (toim.) 2019. Kansalliset laaturekisterit sosiaali- ja terveydenhuollossa - Toimintamalli, organisointi ja rahoitus. Terveyden ja hyvinvoinninlaitos (THL). Terveyden ja hyvinvoinnin laitoksen raportti 16/2019. Helsinki. <http://urn.fi/URN:ISBN:978-952-343-420-2>
- Kirk Roger 2013. *Experimental design. Procedures for the behavioral sciences*. 4. painos. SAGE Publications, Los Angeles, California.
- Kivipelto Minna, Blomgren Sanna, Karjalainen Pekka & Saikkonen Paula 2013. *Vaikuttavaa aikuissosiaalityötä – arviointimalleista mittareihin. Tutkimus- ja kehittämishankkeen loppuraportti*. Terveyden ja hyvinvoinnin laitoksen raportti 8, Tampere. <http://urn.fi/URN:ISBN:978-952-245-864-3>
- Komulainen Jorma, Vuorela Piia & Malmivaara Antti 2014. Satunnaistetun kontrolloidun tutkimuksen periaatteita ja sudenkuoppia. Tutkimustiedon kriittinen arviointi. *Lääketieteellinen Aikakauskirja Duodecim* 130(14), 1439–1444.
- Lammintakanen Johanna, Rissanen Sari, Peronmaa-Hanska Erja, Joensuu Maarit & Ruottu Tiina 2016. Johtaminen ja kehittäminen sosiaali- ja perhehoidossa. Monialaisen ja ammattiryhmäkohtaisen toiminnan käytännöt ja rakenteet. Sosiaali- ja terveysministeriön raportteja ja muistioita 68, Helsinki. <http://urn.fi/URN:ISBN:978-952-00-3851-9>
- L1326/2010. *Terveydenhuoltolaki*. <https://finlex.fi/fi/laki/ajantasa/2010/20101326>
- L1301/2014. *Sosiaalihuoltolaki*. <https://www.finlex.fi/fi/laki/ajantasa/2014/20141301>
- L552/2019. *Laki sosiaali- ja terveystietojen toissijaisesta käytöstä*. <https://www.finlex.fi/fi/laki/alkup/2019/20190552>
- Levasseur Pierre 2019. Implementing a regression discontinuity design to explore the heterogeneous effects of obesity on labour income: the case of Mexico. *Journal of Public Health* 27(1), 89–101.

- Loi Massimo & Rodrigues Margarida 2012. *A note on the impact evaluation of public policies: the counterfactual analysis*, Munich Personal RePEc Archive Paper 42444, Germany: University Library of Munich. <https://mpra.ub.uni-muenchen.de/42444/>
- Lundqvist Martina, Alwin Jenny, Henriksson Martin, Husberg Magnus, Carlsson Per & Ekdahl Anne 2018. Cost-effectiveness of comprehensive geriatric assessment at an ambulatory geriatric unit based on the Age-FIT trial. *BioMed Central Geriatrics* 18(1), 32.
- Malmivaara Antti & Komulainen Jorma 2014. Tutkimustiedon kriittinen arviointi: Luotettavaa vaikuttavuustietoa järjestelmällisistä katsauksista. *Lääketieteellinen Aikakauskirja Duodecim* 130(16), 1635–1641.
- McCoy Erik C. 2017. Understanding the Intention-to-treat Principle in Randomized Controlled Trials. *West Journal of Emergency Medicine* 18(6), 1075–1078.
- McDavid James C., Huse Irene & Hawthorn Laura R.L. 2013. *Program evaluation and performance measurement. An introduction to practice*. 2. painos. SAGE Publications, Los Angeles.
- Meltzer David & Smith Peter 2011. Theoretical issues relevant to the economic evaluation of health technologies. Teoksessa Pauly Mark, McGuire Thomas & Barros Pedro (toim.) *Handbook of Health Economics*, Vol 2. Elsevier, Nort Holland, 433–470.
- Monahan Kathryn, Lee Joanna M. & Steinberg Laurence 2011. Revisiting the impact of part-time work on adolescent adjustment: distinguishing between selection and socialization using propensity score matching. *Child Development* 82(1), 92–112.
- Ngandu Tiia, Lehtisalo Jenni, Solomon Alina, Levälahti Esko, Ahtiluoto Satu, Antikainen Riitta, Bäckman Lars, Hänninen Tuomo, Jula Antti, Laatikainen Tiina, Lindström Jaana, Mangialasche Francesca, Paajanen Teemu, Pajala Satu, Peltonen Markku, Rauramaa Rainer, Stigsdotter-Neely Anna, Strandberg Timo, Tuomilehto Jaakko, Soininen Hilikka & Kivipelto Miia 2015. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *Lancet* 385(9984), 2255–2263.
- Netten Ann, Burge Peter, Malley Juliette, Potoglou Demetris, Towers Ann-Marie, Brazier John, Flynn Terry, Forder Julien & Wall Berry 2012. Outcomes of social care for adults: developing a preference-weighted measure. *Health Technology Assessment* 16(16), 1–166.
- O’Neill Donal, McGilloway Sinéad, Donnelly Michael, Bywater Tracey, Kelly Paul 2013. A cost-effectiveness analysis of the Incredible Years parenting programme in reducing childhood health inequalities. *The European Journal of Health Economics* 14(1), 85–94.

- Paasio Petteri 2006. Yleinen ja erityinen viitekehys arvioinnista. *Hallinnon tutkimus* 25(3), Arvioinnin teemanumero, 92–107.
- Pawson Ray & Tilley Nick 1997. *Realistic evaluation*. SAGE Publications, London.
- Pekkarinen Tuomas 2006. Toimenpiteiden vaikutusten arviointi empiirisessä mikrotaloustieteessä. *Kansantaloudellinen aikakauskirja* 3, 307–320.
- Prieto Luis & Sacristán José A 2003. Problems and solutions in calculating quality-adjusted life years (QALYs). *Health and Quality of Life Outcomes* 1(1), 80.
- Rajavaara Marketta 2007. *Vaikuttavuusyhteiskunta. Sosiaalisten olojen arvostelusta vaikutusten todentamiseen*. Väitöskirja. Sosiaali- ja terveysturvan tutkimuksia 84. Kelan tutkimusosasto, Helsinki.
- Rannanheimo Piia & Jauhonen Hanna-Mari 2018. *Mihin reaaliaikaisen maailman dataa tarvitaan? Näkökulmana lääkkeiden ja lääkinnällisten laitteiden valvonta, arviointi (HTA) ja kansallinen ohjaus*. Sosiaali- ja terveystieteiden tutkimuskeskuksen raportteja ja muistioita 44. Sosiaali- ja terveystieteiden tutkimuskeskus, Helsinki.
- Robson Colin 2001. *Käytännön arvioinnin perusteet*. Opas evaluaation tekijöille ja tilaajille. Suom. työryhmä: Tuija Lindqvist, Maisa Maaniittu, Eila Niemi, Petteri Paasio & Leea Paija. Tammi, Helsinki.
- Rosenbaum Paul R. & Rubin Donald B. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70(1), 41–55.
- Rosenbaum Paul R. & Rubin Donald B. 1984. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 79(387), 516–524.
- Rossi Peter H., Lipsey Mark W. & Freeman Howard E. 2004. *Evaluation: A systematic approach*. 7. painos. Sage, Thousand Oaks Calif.
- Rubin Allen & Babbie Earle 2008. *Research Methods for Social Work*. Australia: Wadsworth/Thomson Learning, Belmont, CA.
- Rutter Michael 1994. Beyond longitudinal data: causes, consequences, changes, and continuity. *Journal of consulting and clinical psychology* 62(5), 928–940.
- Saarni Lea 2009. *Kontrolloitu interventiotutkimus koulupisteiden vaikutuksista koululaisten tuki- ja liikuntaelinten terveyteen*. Väitöskirja. Acta Electronica Universitatis Tamperensis 817. Tampere University Press, Tampere.
- Sackett David L. 2011. Explanatory and pragmatic clinical trials: a primer and application to a recent asthma trial. *Polskie Archiwum Medycyny Wewnetrznej* 121(7–8), 259–263.
- Schulpher Mark 2010. The role and estimation of productivity costs in economic evaluation. Teoksessa Drummond Michael, McGuire Alistair (toim.) *Economic evaluation in health care. Merging theory with practice*. Oxford University Press, Oxford, 94–112.

*Sosiaali- ja terveyspalvelujen vaikuttavuuden ja
kustannusvaikuttavuuden tutkiminen*

- Sintonen Harri 2001. The 15D instrument of health-related quality of life: properties and applications. *Annals of Medicine* 33(5), 328–336.
- Sintonen Harri & Pekurinen Markku 2006. *Terveystaloustiede*. WSOY, Porvoo.
- Sosiaalihuollon lainsäädännön uudistamistyöryhmä 2012. *Sosiaalihuollon lainsäädännön uudistamistyöryhmän loppuraportti*. Sosiaali- ja terveysministeriön raportteja ja muistioita 21, Helsinki.
- Steinberg Laurence, Fegley Suzanne & Dornbusch Sanford M. 1993. Negative Impact of Part-Time Work on Adolescent Adjustment: Evidence From a Longitudinal Study. *Developmental psychology*, 29(2), 171–180.
- Sund Reijo 2003. Utilisation of administrative register using scientific knowledge discovery. *Intelligent Data Analysis* 7(6), 501–519.
- Sund Reijo, Nylander Olli & Palonen Tuula 2004. Raa’asta rekisteriaineistosta terveyspoliittisesti relevanttiin informaatioon. *Yhteiskuntapolitiikka* 69(4), 372–379.
- Thistlethwaite Donald L. & Campbell Donald T 1960. Regression-discontinuity Analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology* 51(6), 309–317.
- TOIMIA-tietokanta 2020. Terveyden- ja hyvinvoinnin laitos. <https://www.terveysportti.fi/dtk/tmi/koti>
- Torkki Paulus, Leskelä Riikka-Leena, Linna Miika, Torvinen Anna, Klemola Katja, Sinivuori Kari, Larsio Antti & Hörhammer Iris 2017. *Ehdotus sosiaali- ja terveyspalveluiden uudeksi kansalliseksi mittaristoksi*. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 36. <http://urn.fi/URN:ISBN:978-952-287-383-5>
- Töttö Pertti. 2004. *Syvällistä ja pinnallista. Teoria, empiria ja kausaalisuus sosiaalitutkimuksessa*. Vastapaino, Tampere.
- Valentijn Pim P. 2015. *Rainbow of chaos: A study into the theory and practice of integrated primary care*. Dissertation. Tilburg University, The Netherlands. https://pure.uvt.nl/ws/portalfiles/portal/9272073/Valentijn_Rainbow_16_12_2015.pdf
- Van den Berg Bernard, Bleichrodt Han & Eeckhoudt Louis 2005. The economic value of informal care: a study of informal caregivers’ and patients’ willingness to pay and willingness to accept for informal care. *Health Economics* 14(4), 363–376.
- Vedung Evert 1997. *Public policy and program evaluation*. Transaction Publishers, New Brunswick (U.S.A).
- Venkataramani Theendar S., Bor Jacob & Jena Anupam B. 2016. Regression discontinuity designs in healthcare research. *British Medical Journal* 352(8050), i1216.
- Virtanen Petri 2007. *Arviointi: arviointitiedon luonne, tuottaminen ja hyödyntäminen*. Edita, Helsinki.

- Vuorela Piia, Malmivaara Antti, Komulainen Jorma & Jousilahti Pekka 2014. Tutkimustiedon kriittinen arviointi: Miten arvioin ja hyödynnän havainnoivan tutkimuksen tuottamaa tietoa? *Lääketieteellinen Aikakauskirja Duodecim* 130(15), 1545–1550.
- Väisänen Antti, Linnosmaa Ismo, Kestilä Laura, Paananen Reija & Gissler Mika 2014. Sijaishuollon vaikutus koulutukseen ja rikollisuuteen, Teoksessa Mäklin S (toim.) *Terveystaloustiede 2014*. Terveyden ja hyvinvoinnin laitoksen Työpäpaperi 3, 39–45. Terveyden ja hyvinvoinnin laitos, Tampere.
- Väyrynen Riikka, Ailio Erja, Arajärvi Miina & Forsell Martta 2020. Sosiaalihuollon tiedonkeruu-uudistuksen kehittämispolku. Versio 1.0. Terveyden ja Hyvinvoinnin laitos. Saatavissa: <http://urn.fi/URN:NBN:fi-fe2020051229433>
- Walton Marc, Powers John, Hobart Jeremy, Patrick Donald, Marquis Patrick, Vamvakas Spiros, Isaac Maria, Molsen Elizabeth, Cano Stefan & Burke Laurie 2015. Clinical Outcome Assessments: Conceptual Foundation—Report of the ISPOR Clinical Outcomes Assessment – Emerging Good Practices for Outcomes Research Task Force. *Value in Health* 18(6), 741–752.
- Weinstein Milton C., Torrance George & McGuire Alistair 2009. QALYs: The Basics. *Value in Health* 12(1), 5–9.
- Whitehead Sarah J. & Ali Shehzad 2010. Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin* 96(1), 5–21.
- WHO *Quality of Life*. <https://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/>
- Williams Alan 1996. QALYs and ethics: A health economist's perspective. *Social Science & Medicine* 43(12), 1795–1804.
- Williams Hywel, Burden-Teh Esther & Nunn Andrew 2015. What is a Pragmatic clinical trial? *Journal of Investigative Dermatology* 135(6), 1–3.
- Wing Coady, Simon Kosali & Bello-Gomez Ricardo A. 2018. Designing Difference in Difference Studies: Best Practices for Public Health Policy Research. *Annual Review of Public Health* 1(39), 453–469.