



Verkkokatsauksia 40/2014

Sisällys

Tiivistelmä	1
1 Johdanto	1
2 Aggredin työmuoto ja tavoitteet	2
3 Aineisto ja menetelmät	2
4 Tulokset	5
5 Pohdintaa	8
Lähteet	10

Mikko Aaltonen & Ville Hinkkanen

Katuväkivaltatyön vaikuttavuus: HelsinkiMission Aggredi-ohjelman arviointi

Tiivistelmä

- HelsinkiMission Aggredi-ohjelman tavoitteena on vähentää 18–39-vuotiaiden katuväkivaltaan syyllistyneiden henkilöiden väkivaltarikosten uusimista. Ohjelman asiakkailla on keskimäärin varsin raskas rikostausta.
- Tässä tutkimuksessa arvioitiin Aggredin asiakkaiden rikosten uusimista sekä ohjelman eri asiakasryhmissä että suhteessa iän, sukupuolen ja rikostaustan perusteella kaltaistettuun verrokkiryhmään.
- Ohjelman tulokset ovat hyviä sen läpikäyneessä ryhmässä: he uusivat rikoksia huomattavasti muita asiakasryhmiä harvemmin. Läpikäyneistä 25 prosenttia uusi väkivaltarikoksen seurantajaksolla.
- Ohjelman keskeyttäneiden uusimisalttius oli sen sijaan korkea. Kaksi kolmasosaa heistä uusi väkivaltarikoksen seuranta-aikana.
- Kun kaikkia Aggredi-ohjelman aloittaneita verrattiin kaltaistettuun verrokkiryhmään, eroa rikosten uusimisessa ei havaittu. Tämä johtuu keskeyttäneiden korkeasta uusintarikollisuudesta, mikä on ohjelman keskeinen haaste.
- Tutkimuksen rajoitteena oli se, että rekisteripohjainen vertailuryhmä ei todennäköisesti ollut lähtökohdiltaan yhtä väkivalta-altis kuin Aggredin asiakasryhmä. Siksi havaintoja arvioitaessa on syytä painottaa ohjelman suorittaneiden alhaista uusintarikollisuutta. Tältä osin tuloksia voidaan pitää varsin lupaavina.
- Kokonaisuutena tarkastellen arviointi viittaa siihen, että Aggredi-ohjelmaa kannattaa jatkaa ja kehittää edelleen.
- Jatkossa kansalaisjärjestöjen toteuttamia rikosentorjuntaohjelmia tulisi arvioida satunnaistettujen koeasetelmien avulla.

1 Johdanto

Aggredi-ohjelma (aikaisemmalta nimeltään ”Aikalisä” vuosina 2006–2009) aloitettiin vuonna 2006 HelsinkiMissio ry:n toimesta Raha-automaattiyhdistyksen rahoituksella. Ohjelman keskeisenä tavoitteena on vähentää katuväkivallasta epäiltyjen ja tuomittujen 18–39-vuotiaiden väkivaltaista käyttäytymistä ja väkivaltarikosten uusimista. Aggredi ja Oikeuspoliittinen tutkimuslaitos tekivät sopimuksen Aggredin vaikuttavuuden arvioinnista vuonna 2009. Tämän tutkimuksen tavoitteena on arvioida sitä, vähentääkö Aggredi-ohjelmaan osallistuminen myöhempää väkivaltarikollisuutta. Tutkimuksen rahoittajina ovat toimineet oikeusministeriön kriminaalipoliittinen osasto ja rikosentorjuntaneuvosto.

Tässä tutkimuksessa tutkimushenkilöiden väkivaltarikollisuuden määrää hoidon aloittamisen jälkeen verrataan iän, sukupuolen ja rikostaustan perusteella kaltaistettuun vertailuryhmään, jonka muodostamisessa käytettiin Oikeuspoliittisen tutkimuslaitoksen Rikosten ja seuraamusten tutkimusrekisterin tietoja henkilöiden aikaisemmasta rikollisuudesta ja rikosseuraamuksista. Osittain vastaavanlaista menetelmää on käytetty muun muassa arvioidessa OMA-ohjelman vaikutusta uusintarikollisuuteen (Konttila & Tyni 2011) sekä tutkittaessa ehdottomaan vankeuteen tai yhdyskuntapalveluun rattijuopumuksesta tuomittujen rikosten uusimista (Sirén

Yhteystiedot

Oikeuspoliittinen tutkimuslaitos
Pitkäsillanranta 3 A, 5. krs
PL 444, FI-00531 Helsinki
Puh. 029 5665300
Fax. 029 5665365

2012). Tutkimushankkeen oheistavoitteena oli arvioida tällaisen menetelmän soveltuvuutta rikosentorjunta-hankkeiden vaikuttavuusarviointiin yleisemminkin.

2 Aggredin työmuoto ja tavoitteet

Seuraavassa kuvataan lyhyesti Aggredi-ohjelman sisältö olennaisilta osiltaan. Aggredin hoitoideologiaa ja -menetelmää on kuvattu tarkemmin Markus Laineen ja Martti Tuomisen (2012) laadullisessa arvioinnissa, Aggredi-julkaisussa (Kekki & Salakka 2013) sekä Aggredin internet-sivuilla (<http://www.helsinkimissio.fi/?sid=474>).

Helsingin Sörnäisissä toimivan Aggredin asiakkaat ohjautuvat toimintaan pääasiassa vankiloiden, yhdyskuntaseuraamustoimistojen, poliisin ja lastensuojelun kautta. Nämä tahot valikoivat omista asiakkaistaan henkilöt, joille ohjelma heidän käsityksensä mukaan voisi sopia (Laine & Tuominen 2012). Aggredin aloittamisen yhteydessä arvioidaan ensisijaisesti sitä, onko henkilö valmis työskentelemään elämäntilanteensa muuttamiseksi. Asiakkaan täytyy olla valmis puhumaan asioistaan. Mikäli näin ei ole, asiakasprosessia ei aloiteta. Näin ollen ohjelman valikoituminen tapahtuu kahdessa vaiheessa (Laine & Tuominen 2012). Toisaalta on huomionarvoista, että Aggrediin osallistuminen on täysin vastikkeetonta, eikä osallistumiselle aseteta muuten tiukkoja ehtoja. Esimerkiksi päihitteettömyyttä ei vaadita, eikä väkivaltarikoksen uusiminen ohjelman aloittamisen jälkeen merkitse asiakassuhteen automaattista loppumista. Aggredi ei siis pyri nollatoleranssiin rikosten suhteen. Aggredi pyrkii myös erottautumaan rikosoikeusjärjestelmästä, ja haluaa näyttäytyä itsenäisenä sosiaalipalveluna rikosseuramusalan jatko- tai jälkihuollon sijaan. Asiakasprosessin alussa korostetaan Aggredin puolueetonta asemaa (Kekki & Salakka 2013).

Asiakastapaamiset järjestetään keskimäärin viikon tai kahden viikon välein, ja käyntikerran pituus vaihtelee yhdestä kahteen tuntiin. Mikäli asiakas suorittaa ehdotonta vankeusrangaistusta ohjelman alkaessa, asiakastyö aloitetaan noin kaksi kuukautta ennen vapautumista tai koevapautta. Asiakasprosessit eivät ole määrämuotoisia, vaan muotoutuvat asiakkaan tarpeiden mukaisesti. Asiakkaat tapaavat kaksi työntekijää yksin, vertaisryhmäkeskusteluja ei järjestä. Asiakastyön muotoja ovat mm. ”motivoiva keskustelu, hoidollinen keskustelu, toimintakyvyn palauttamiseen tähtäävä kriisikeskustelu tai arkielämän organisointi” (Kekki & Salakka 2013, 21). Asiakastyö nojaa ajattelutapaan ja tutkimustraditioon, jossa toimijoiden tilanteenmäärittelyjen ja tulkintojen sekä niissä tapahtuvien muutosten katsotaan vaikuttavan heidän myöhempään käyttäytymiseensä. Keskusteluissa pyritään vuoropuheluun asiakkaan kanssa siinä kontekstissa, jossa hän itse kokee elävänsä, ja pyritään välttämään tilannetta, jossa ainoastaan ympäröivän yhteiskunnan ja työntekijöiden arvot määrittäisivät asiakkaan elämäntilanteen

tulkintaa (Kekki & Salakka 2013). Tavoitteena on, että ”uusien näkökulmien kautta asiakkaalle alkaa muodostua uutta identiteettiä, mikä ei ole sidoksissa siihen kontekstiin, jossa asiakas normaalisti käyttää väkivaltaa” (Kekki & Salakka 2013, 23). Tällä pyritään siihen, että uudet identiteetit ja roolit syrjäyttäisivät väkivaltarikollisen roolin. Laadullisessa arvioinnissa Aggredin toimintatapojen kuvataan muistuttavan ulkopuolelta katsottuna ”kognitiivista psykoterapiaa, jonka kantava ajatus on opittujen huonojen käyttäytymistä ohjaavien mielikuvien (automaattisten ajatusten) kyseenalaistaminen ja niiden muuttaminen vähemmän automaattisiksi” (Laine & Tuominen 2012, 54).

Aggredin tärkeimpänä tavoitteena on vähentää asiakkaidensa väkivaltaista käyttäytymistä, ja lisäksi myötävaikuttaa elämäntilanteen paranemiseen ja hyvinvoinnin lisääntymiseen. Rikollisen elämäntyylin vähentymisen ja väkivaltarikollisuuden loppumisen lisäksi merkinä edistyksistä on Aggredin kriteerien mukaan se, jos henkilö pystyy arvioimaan tekojensa seurauksia ja suhdetta rikolliseen elämäntapaan paremmin kuin aikaisemmin (Kekki & Salakka 2013). Sitoutumisen asetta kuvaavat mm. käyntikertojen määrä ja asiakkuuden kesto. Aggredi pyrkii neuvomaan ja jatko-ohjaamaan asiakkaitaan muiden palveluiden ääreen, jotta he saisivat apua esimerkiksi asunnon, toimeentulon tai päihdeongelman kanssa. Aggredin asiakkuus päättyy tyypillisesti joko työntekijän ja asiakkaan yhteisellä sopimuksella tai asiakkaan yksipuolisella päätöksellä. Asiakkaalla on aina mahdollisuus ottaa yhteyttä uudelleen tai palata takaisin asiakkaaksi. Osa asiakasprosesseista päättyy asiakkaan palatessa takaisin vankilaan tai tilanteessa, joissa päihneiden ongelmakäyttö ja rikollisuus jatkuvat intensiivisinä (Salakka & Kekki 2013).

3 Aineisto ja menetelmät

Vaikuttavuusarvioinnin keskeisenä tavoitteena on löytää hoitoon osallistuneille henkilöille uskottava verrokkiryhmä, jonka avulla voidaan muodostaa arvio siitä, mitä hoitoon osallistuneille henkilöille olisi tapahtunut ilman hoitoon osallistumista. Satunnaistetussa koeyhteisessä hoito- ja verrokkiryhmien keskimääräinen samankaltaisuus taataan siten, että ohjelmaan soveltuvat henkilöt arvotaan näihin kahteen (tai useampaan) ryhmään ennen hoidon alkua. Henkilöiden määrän ollessa riittävän suuri arpominen takaa sen, että hoito- ja verrokkiryhmät ovat keskimäärin samanlaiset kaikkien havaittujen ja havaitsemattomien muuttujien suhteen. Tavoitteena on siis valikoitumisharhan (selection bias) minimointi. Koska Aggrediin osallistumisen satunnaistaminen ei ollut mahdollista, verrokkeina toimivat henkilöt, jotka ovat iän, sukupuolen ja rikostaustan perusteella keskimäärin samanlaisia kuin Aggrediin osallistuneet henkilöt.

Vaikuttavuusarvioinnissa tunnetaan kaksi lähestymistapaa liittyen hoitoryhmän määrittelyyn. Tässä

tutkimuksessa tuloksia esitetään molempien alla kuvattujen määrittelyjen mukaisesti.

Hoitoaikkeen mukainen analyysi (Intention-to-treat analysis [ITT], Sarna 2012). Tähän laajimpaan hoitoryhmämäärittelyyn kuuluvat tässä yhteydessä kaikki henkilöt, joille tarjottiin mahdollisuutta osallistua Aggredi-ohjelmaan (n=141). Lähtöoletuksena oli, että erityisen väkivalta-alttiita henkilöitä, joiden uusimiskorkeus on korkea, olisi vaikeampi saada kiinnittymään ohjelmaan. ITT-asetelmassa hoidon vaikutusta ei siis arvioida ainoastaan hoidon onnistuneesti suorittaneessa ryhmässä, vaan koko siinä ryhmässä, jolla olisi ollut mahdollisuus osallistua hoitoon. Näin toimitaan pääasiassa kahdesta syystä: Ensisijaisesti siksi, että tutkimus- ja verrokkiryhmät pysyisivät vertailukelpoisina, eikä tutkimusryhmän motivoituneita henkilöitä verrattaisi kaltaistetun verrokkiryhmän motivoituneisiin ja ei-motivoituneisiin verrokkeihin. Satunnaistetussa asetelmassa hoitoaikkeen mukainen analyysi takaa alkuperäisen satunnaistamisen vahvuuksien (tutkimus- ja vertailuryhmä ovat keskimäärin samanlaiset kaikilla havaituilla ja havaitsemattomilla muuttujilla) säilymisen. Toinen syy on se, että tyypillisesti vaikutusarvioinnissa halutaan arvioida hoidon toimivuutta koko kohdejoukossa, ei ainoastaan ohjelman läpikäynnissä osaryhmässä. Pelkästään ohjelman suorittaneiden perusteella tehty arvio ohjelman tehosta saattaisi helposti johtaa ylioptimistisiin odotuksiin siitä, millainen teho ohjelman laajentamisella esimerkiksi toiselle paikkakunnalle olisi. (Gupta 2011)

Hoitotehoanalyysi (Efficacy analysis, Sarna 2012). Hoitotehoanalyysin kohteena ovat ne henkilöt, jotka todella osallistuivat ohjelmaan. Aggredin osalta hoitotehoanalyysi voi kohdistua kahteen eri ryhmään: niihin, jotka aloittivat ohjelman (n=97), tai niihin, jotka suorittivat ohjelman onnistuneesti loppuun (n=24). Jos hoitoaikkeen mukaista analyysia ja hoitotehoanalyysia verrataan, tarjoaa ensimmäinen vaihtoehto lähtökohtaisesti luotettavamman arvion ohjelman kausaalista vaikutuksesta. ITT-asetelman heikkoutena on puolestaan se, että estimaatti hoidon vaikutuksesta on lähtökohtaisesti konservatiivinen, ja todellinen ero tutkimusryhmän ja verrokkien välillä jää helposti havaitsematta. Tämä ongelma on erityisen hankala arvioinneissa, joissa valtaosa tutkimusryhmään päätyneistä ei suostu va-

paaehtoiseen ohjelmaan tai jättää sen kesken. Näin on myös Aggredin tapauksessa. Kolmasosa niistä (n=44), joille ohjelmaa tarjottiin, ei päätenyt ohjelmaan lainkaan (todellisuudessa tämä osuus on vielä suurempi, sillä täydellistä listaa henkilöistä, joiden kanssa Aggredista oli keskusteltu, ei ole olemassa). Ohjelman aloittaneesta 97 henkilöstä vain neljäsosa suoritti sen onnistuneesti loppuun siten, että asiakasprosessi katsottiin Aggredin toimesta päättyneeksi.

Edellä mainituista syistä johtuen ohjelman toimitusta arvioidaan tässä tutkimuksessa kolmessa eri osajoukossa: ITT-ryhmässä (n=141), hoidon aloittaneiden ryhmässä (n=97) ja ohjelman onnistuneesti läpikäyneiden ryhmässä (n=24). Tutkimusryhmäksi kutsutaan seuraavassa sitä joukkoa, joka aloitti Aggredi-ohjelman. Tämä ryhmä on aineistona analyysin siinä osassa, jossa tarkastellaan rikosuusimista Aggredi-ohjelman sisäisten tekijöiden suhteen. Tällaisia muuttujia ovat hoidon päättymisen syy, käyntien lukumäärä, ohjaava taho, asiakkaan väkivaltaprofiili sekä asiakkaan ikä. Verrokkiryhmävertailu tehdään puolestaan kaikissa kolmessa ryhmässä.

Taulukossa 1 on raportoitu ikä, sukupuoli ja rikostausta kolmessa edellä mainitussa ryhmässä. Näiden tietojen perusteella vaikuttaa siltä, että ohjelman aloittaneilla ja läpikäyneillä on keskimäärin varsin samanlainen rikostausta sekä vakivaltarikosten että kaikkien rikosten määrällä mitattuna. Lähes kaikki ohjelmaan osallistuneet olivat miehiä, ja ohjelman läpikäyneet hieman koko tutkimusryhmää vanhempia. Kaiken kaikkiaan ohjelman tavoittama ryhmä on varsin rikosaktiivista. Kaikkien tuomioiden ja rangaistuspäättämisen vuotuinen keskiarvo on koko aineistossa hieman alle seitsemän. Hoidon alkamishetkellä 56 prosenttia henkilöistä suoritti vankeusrangaistusta tai oli tutkintavankeudessa. Tutkimusryhmän, eli hoidon aloittaneiden joukosta noin 80 prosenttia tuli Aggrediin vankiloiden tai yhdyskuntaseuraamustoimistojen kautta. Seuranta-ajan pituudessa ei ole kovin suurta eroa, ohjelman läpikäyneiden keskimääräinen uuden rikollisuuden seuranta-aika on noin kolme kuukautta koko tutkimusryhmää lyhyempi. Lähtökohtainen oletus siitä, että ohjelman keskeyttäneet olisivat lähtökohtaisesti väkivalta-alttiimpia kuin sen läpi käyvät ei siis pidä paikkaansa ainakaan rikostaustalla mitattuna.

Taulukko 1 Tutkimusaineiston taustamuuttujat (keskiarvo tai prosenttiosuus)

	ITT (n=141)	Tutkimusryhmä (n=97)	Läpikäyneet (n=24)
Ikä ohjelman alussa	23,9	24,4	25,2
Mies	98 %	98 %	100 %
Väkivaltarikosten määrä 3 vuotta ennen	3,7	3,7	3,9
Kaikkien rikosten määrä 3 vuotta ennen	20,5	20,6	18,5
Päiviä vankeudessa 3 vuotta ennen	319	360	384
Vankilassa alkupäivänä	45 %	56 %	63 %
Seuranta-aika (pv)	1 060	1 097	998

Verrokkien poiminta ja kaltaistaminen. Verrokkiryhmä muodostettiin Oikeuspoliittisen tutkimuslaitoksen Rikosten ja seuraamusten tutkimusrekisterin tietojen perusteella. Ensimmäisessä vaiheessa jokaiselle ITT-ryhmän henkilölle poimittiin maksimissaan kymmenen samanikäistä (+/- 1 vuosi) ja samaa sukupuolta olevaa verrokkia, joiden rikostausta rikosseuraamusten valossa kolme vuotta ennen seurannan alkua muistutti tutkimushenkilöiden rikostausta. Verrokkien henkilöiden seurannan alku määritettiin päivälleen samaksi kuin tutkimushenkilöiden. Poiminnassa löytyi yhteensä 1 112 verrokkia (1 041 erillistä henkilöä), eli noin 7,7 verrokkia per tutkimushenkilö. Koska Aggredi keskittyy nimenomaan katuväkivallasta epäiltyihin henkilöihin, verrokkiryhmän henkilöt, joilla oli taustallaan joko seksuaalirikoksia tai perheväkivaltaa tiputettiin pois aineistosta. Samoin poistettiin sellaiset verrokkit, jotka olivat vankeudessa koko seuranta-ajan.

Ennen hoidon alkua mitattuja kaltaistusmuuttujia olivat iän ja sukupuolen lisäksi väkivaltarikosten kokonaisuus, törkeään pahoinpitelyyn syyllistyminen (kyllä/ei), ehdottomien vankeustuomioiden lukumäärä sekä vankeuspäivien lukumäärä. Nämä muuttujat kuvaavat kolme vuotta ennen seurannan alkua. Lisäksi käytetään tietoa poliisin tietoon tulleista väkivaltarikoksista vuotta ennen hoidon alkua, sekä tietoa siitä, suorittiko henkilö ehdotonta vankeusrangaistusta seurannan alkupäivänä.

Vaikka verrokkien poiminnan ensimmäisessä vaiheessa käytettiin jo suhteellisen tiukkoja kriteereitä rikostaustan vastaavuuden suhteen, tutkimus- ja verrokkiryhmien tarkempi vertaaminen edellä mainittujen taustamuuttujien suhteen osoitti, että keskimääräinen rikostausta verrokkiryhmässä oli huomattavasti ITT-ryhmää kevyempi. Seuraavassa vaiheessa pyrittiin löytämään jokaiselle tutkimushenkilölle vielä tarkemmat verrokkit kaltaistamismenetelmiä (matching) käyttämällä. Näillä menetelmillä pyritään saavuttamaan satunnaistetun koeasetelman hyödyt havainnoivassa aineistossa (Austin 2011). Kaltaistamismenetelmien tavoitteena on löytää hoitoryhmälle taustatekijöiden valossa mahdollisimman samankaltaisia verrokkien henkilöitä, jotka eivät osallistuneet hoitoon. Nämä taustatekijät ovat tyypillisesti sellaisia muuttujia, joilla on yhteys sekä ohjelmaan osallistumisen todennäköisyyteen että varsinaiseen vastetapahtumaan, tässä tapauksessa uusintarikollisuuteen.

Kun tutkija on tehnyt valinnan taustamuuttujista, joita kaltaistuksessa käytetään, seuraavaksi on valittava menetelmä, jolla henkilöiden samankaltaisuutta näiden muuttujien suhteen arvioidaan. Kun kaltaistusmuuttujia on useita, verrokkien potentiaalisesta joukosta on yleensä vaikea löytää henkilöitä, jotka saisivat täsmälleen samat arvot kuin tutkimusryhmän henkilö kaikilla käytetyillä muuttujilla (ns. curse of dimensionality, ks. Austin 2011). Tästä syystä sataprosenttisen tarkka kaltaistus (exact matching) ei yleensä onnistu. Tyypillisin keino kiertää tämä ongelma on käyttää kal-

taistamisen perusteena ns. propensiteettipistemäärää (propensity score). Propensiteettipistemäärä luodaan siten, että esimerkiksi logistisen regressioanalyysin avulla selitetään hoitoon osallistumista kuvaavaa binääristä muuttujaa kaikilla valituilla kaltaistamis- muuttujilla, ja sen jälkeen tehdään malliin perustuva ennuste kunkin henkilön todennäköisyydestä osallistua hoitoon (joko lineaarisena ennusteena tai ennustettuna todennäköisyytenä). Näin luodun propensiteettipistemäärän hyödyllisyys perustuu sen tasapainottavaan ominaisuuteen (balancing score, Austin 2011), eli siihen, että samansuuruisen propensiteettipistemäärän saavien tutkimus- ja verrokkien henkilöryhmien jakauma alkuperäisillä kaltaistusmuuttujilla on myös samankaltainen (Rosenbaum 2010).

Varsinaisten verrokkien valinnassa on lisäksi useampia vaihtoehtoja: voidaan esimerkiksi käyttää pistemäärältään lähintä henkilöä (nearest neighbor), useampaa lähintä naapuria, samaa verrokkia useampaan kertaan, rajata tutkimushenkilöiden ja verrokkien pistemäärän maksimietäisyyttä tai yhdistää propensiteettipistemäärän käyttö tarkkaan kaltaistukseen esimerkiksi iän ja sukupuolen suhteen (Austin 2011). Vaihtoehtoisena menetelmänä propensiteettipistemääräkaltaistukselle on käytetty monimuuttujakaltaistusta (covariate matching), jossa hoitoryhmän ja verrokkien etäisyyttä taustamuuttujien suhteen arvioidaan ns. Mahalanobis-etäisyyden perusteella (Rubin 1980; Diamond & Sekhon 2013). Käytetään kumpaa tahansa menetelmää, kaltaistuksen perusidea on joka tapauksessa tämän tutkimuksen viitekehäyksessä se, että tutkimushenkilön rikosten uusimista pyritään vertamaan ainoastaan sellaisten verrokkien henkilöiden uusintarikollisuuteen, joilla olisi iän, sukupuolen ja rikostaustan valossa ollut yhtä suuri todennäköisyys osallistua Aggredi-ohjelmaan.

Tässä tutkimuksessa tehty analyysi olisi voitu periaatteessa toteuttaa normaalin regressioanalyysin keinoin, jossa hoitoon osallistumista kuvaava binääri- nen muuttuja olisi vakioitu yhtä aikaa taustamuuttujien kanssa, ja arvioitu eroa rikosten uusimista kuvaavassa vastemuuttujassa tuon binäärisen muuttujan saaman estimaatin ja sen tilastollisen merkitsevyyden avulla. Regressioanalyysin osalta ongelmaksi muodostuu kuitenkin se, ettei siinä lähtökohtaisesti tarkastella sitä, missä määrin taustamuuttujien valossa vertailukelpoisia yksilöitä todella löytyy, ja missä määrin tutkimus- ja verrokkiryhmät onnistutaan tasapainottamaan (Austin 2011). Propensiteettipistemäärään perustuvissa menetelmissä tutkimus- ja verrokkien tasapainon arviointi on sen sijaan ensimmäinen osa analyysiprosessia, joka voidaan erottaa varsinaisesta ohjelman vaikutuksen analyysistä. Propensiteettipistemäärää mallinnettaessa tutkijan on päätettävä tilastollinen malli, joka onnistuu parhaimmalla mahdollisella tavalla tasapainottamaan tutkimus- ja verrokkien ryhmät taustamuuttujien valossa. Parhaan mahdollisen tasapainon löytäminen eri malleja vertaamalla on kuitenkin vaikeaa (Diamond & Sekhon 2013). Tässä tutkimuksessa kal-

taistukseen käytettiin R-ohjelmiston Genetic Matching -algoritmia (Sekhon 2011, ks. myös Skardhamar & Savolainen 2014), joka automatisoi tämän prosessin etsimällä jokaiselle taustamuuttujalle iteratiivisesti optimaalisen painokertoimen, joka maksimoi tutkimus- ja verrokiryhmien tasapainon kaikkien taustamuuttujien suhteen (Sekhon 2011; Diamond & Sekhon 2013). Menetelmä tarjoaa lisäksi hyvän diagnostiikan tasapainotuksen onnistumisen arviointiin¹. Aggredin vaikutusta kuvaavana estimaattina käytetään keskimääräistä hoidon vaikutusta hoidetuille (average treatment effect on the treated), joka kertoo sen, kuinka paljon Aggrediin osallistuneiden rikollisuus poikkeaa suhteessa heidän verrokkeihinsa.

4 Tulokset

Rikosten uusiminen tutkimusryhmässä ja sen osajoukoissa. Kun rikosten uusimista tarkastellaan vain varsinaisessa tutkimusryhmässä, eli siinä joukossa, joka aloitti Aggredi-ohjelman, havaitaan rikollisuuden määrässä varsin merkittäviä eroja tutkimusryhmän eri osajoukoissa. Tärkein ero liittyy siihen, suorittiko henkilö Aggredi-ohjelman loppuun. Ohjelman onnistuneesti läpi käyneiden ryhmässä rikosuusiminen (0,2 väkivaltarikosta ja 1,4 rikosta vuodessa) on selkeästi matalammalla tasolla kuin ohjelman eri syistä keskeyttäneiden ryhmässä. Tämä nähdään sekä taulukon 1 tiedoista että kuvioista 1, joka kuvaa väkivaltarikoksia tehneiden osuuden seuranta-ajan suhteen. Kuviossa 1 pystyviivat kuvaavat sitä, kun tutkimushenkilön yksilöllinen seuranta-aika päättyy vuoden 2012 lopun aikarajoitukseen ilman, että kyseinen henkilö uusii väkivaltarikoksen siihen mennessä. Vaakaviivan siirtymä ylöspäin puolestaan kuvaa sitä hetkeä seurannan alun jälkeen, kun ryhmän henkilö uusii väkivaltarikoksen.

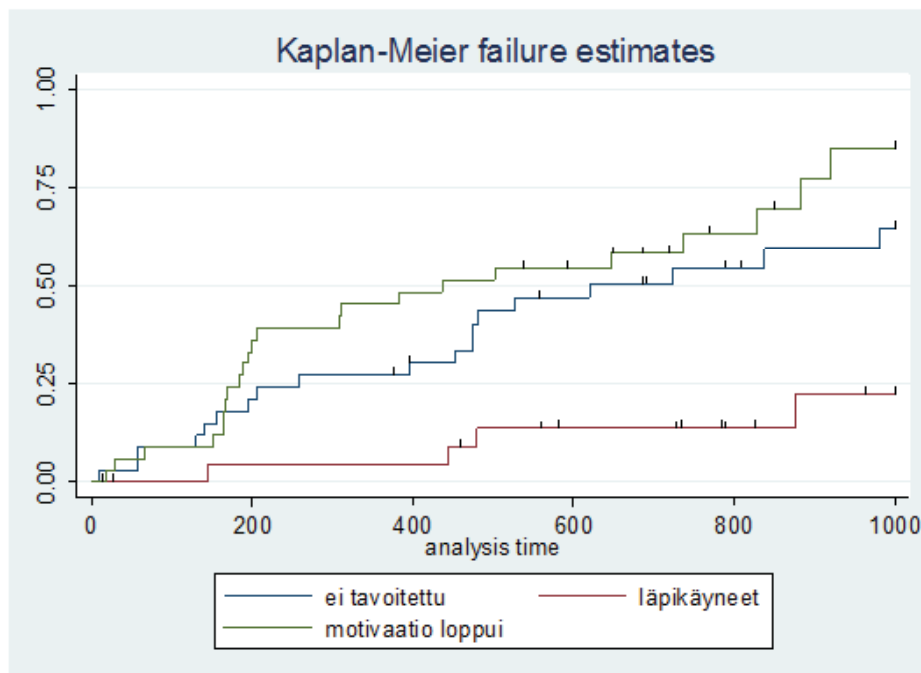
Vastemuuttajat. Rikosten uusimista tutkimus- ja verrokiryhmässä arvioidaan usean eri mittarin avulla. Koska ohjelman tavoitteena on vähentää nimenomaan väkivaltarikoksia, tärkein mittari on väkivaltarikosten määrä seuranta-aikana. Nämä tiedot perustuvat tuomioihin ja rangaistusmääräyksiin. Koska seuranta-aikojen pituudet vaihtelevat henkilöittäin (seuranta loppuu 31.12.2012), rikosten määrä suhteutetaan sekä a) seuranta-ajan pituuteen että b) vapaudessa (ei vankilassa) oltujen päivien lukumäärän. Myös uusineiden prosenttiosuus raportoidaan. Koska yhteiskunnan kannalta myös muun kuin väkivaltarikollisuuden väheneminen olisi suotavaa, tarkastellaan lisäksi kaikkien rikosten määrää seuranta-aikana.

Käyntien lukumäärä on myös yhteydessä uusimistiheyteen: ne henkilöt, jotka ovat käyneet tapaamisessa yli kymmenen kertaa uusivat rikoksensa harvemmin kuin sellaiset henkilöt, jotka kävivät tapaamisissa alle kymmenen kertaa. Tämä ero ei ole kuitenkaan tilastollisesti merkitsevä. On huomioitava, että päätymisen syy ja käyntien lukumäärä ovat vahvassa yhteydessä toisiinsa: ohjelman onnistuneesti päättäneet kävivät tapaamisissa keskimäärin 20 kertaa, kun vastaava määrä muissa ryhmissä on noin viisi kertaa. Kokonaisuutena arvioiden vaikuttaa siis selvältä, että ohjelmaan vahvemmin kiinnittyneet henkilöt uusivat väkivaltarikoksena huomattavasti muita harvemmin. Huomionarvoista on se, että rikollisuuden tasossa kolme vuotta ennen ohjelman alkua ei ole merkittävää eroa päättymisen syyn ja käyntien määrän mukaan.

¹ Vaihtoehtoisina analyysimalleina käytettiin perinteisiä propensiteettipistemäärään perustuvia kaltaistumalleja, joiden sovittamisessa hyödynnettiin Stata 12.1 -ohjelmiston moduuleja psmatch2 (Leuven & Sianesi 2003), nnmatch (Abadie ym. 2004) ja pscore (Becker & Ichino 2002). Eri menetelmät tuottivat pitkälti vastaavia tuloksia.

Taulukko 2 Rikosten uusiminen tutkimusryhmässä ja sen eri osajoukoissa (suluissa yksisuuntaisen varianssianalyysin F-testin p-arvo)

Muuttuja	Väkivaltarikokset			Kaikki rikokset lkm / vuosi
	lkm / vuosi	lkm / vapausvuosi	uusineiden osuus	
Tutkimusryhmä (n=97)	0,53	0,80	0,54	4,51
Päättymisen syy				
Asiakasta ei tavoitettu (n=33)	0,53 (p=0,00)	0,92 (p=0,03)	0,67 (p=0,00)	4,32 (p=0,00)
Motivaatio loppui (n=33)	0,86	1,16	0,70	7,67
Läpikäyneet (n=24)	0,20	0,32	0,25	1,41
Käyntien määrä				
1–4 (n=39)	0,59 (p=0,47)	0,96 (p=0,50)	0,64 (p=0,36)	4,20 (p=0,70)
5–9 (n=20)	0,55	0,81	0,55	5,58
10+ (n=26)	0,38	0,59	0,46	4,09
Ikä				
18–23 (n=50)	0,59 (p=0,59)	0,85 (p=0,65)	0,58 (p=0,34)	5,24 (p=0,40)
24–29 (n=33)	0,50	0,87	0,55	4,26
30–36 (n=14)	0,35	0,53	0,36	2,51
Asiakasprofiili				
Impulsiivinen väkivalta (n=39)	0,37 (p=0,04)	0,53 (p=0,04)	0,44 (p=0,14)	2,43 (p=0,00)
Taloudellista hyötyä tavoitteleva väkivalta (n=45)	0,71	1,07	0,60	6,49
Lähettävä taho				
Vankila (n=42)	0,61 (p=0,35)	0,89 (p=0,43)	0,47 (p=0,20)	6,00 (p=0,13)
Yhdyskuntaseuraamustoimisto (n=35)	0,54	0,88	0,68	3,89
Muu (n=20)	0,32	0,50	0,45	2,49

**Kuvio 1** Aika seuraavaan väkivaltarikokseen ohjelman aloittamisen jälkeen, päättymisen syyn mukaan (päivinä)

Asiakkaan ikä vaikuttaa olevan oletusten mukaisesti yhteydessä rikosten uusimisen intensiteettiin siten, että vanhemmat tutkimushenkilöt uusivat rikoksensa harvemmin, vaikkei ero ikäryhmien välillä olekaan merkittävä. Vanhemmat tutkimushenkilöt kuuluvat lisäksi nuorempia todennäköisemmin ohjelman suorittaneiden ryhmään. Asiakkuusprofiiliin osalta tulokset esitetään vain kahdelle ryhmälle (impulsiivinen väkivalta ja välineellinen/taloudellista hyötyä tavoitteleva väkivalta), sillä muiden ryhmien koot ovat pienet. Näiltä osin vaikuttaa siltä, että hoitoa saanut impulsiivisen väkivallan ryhmä uusii rikoksensa harvemmin, ja erityisesti rikosten kokonaismäärä on huomattavasti toista ryhmää matalampi. Lähehtävän tahon osalta havaitaan, että vankiloiden ja yhdyskuntaseuraamustoimistojen kautta ohjelmaan tulneiden uusimisen taso on väkivallan osalta suunnilleen sama.

Rikosten uusiminen tutkimus- ja verrokkiryhmässä. Edellisessä tarkastelussa havaittiin, että väkivallan uusimisessa on tutkimusryhmän sisällä merkittäviä eroja. Ohjelmaan paremmin kiinnittyneet ja sen loppuun suorittaneet henkilöt syyllistyvät uusiin väkivaltarikoksiin huomattavasti ohjelman keskeyttäneitä harvemmin. On kuitenkin edelleen mahdollista, että ohjelman onnistuneesti suorittaneet henkilöt ovat olleet ylipäättään motivoituneempia muuttamaan käytöstään, ja eri asiakasryhmiä koskeva tulos saattaa ainakin osin heijastaa tällaisia eroja. Rikollisuuden väheneminen olisi siten voinut tapahtua myös ilman ohjelmaan osallistumista. Sen perusteella mitä tiedetään iän ja rikosalttiuden yhteydestä voidaan lisäksi olettaa, että sekä tutkimus- että verrokkiryhmän rikollisuus vähennee todennäköisesti joka tapauksessa ikääntymisen myötä. Tämä on nähtävissä tässäkin aineistossa, ja

siitä syystä saman henkilön rikostasojen vertailusta ennen Aggredin aloittamista ja sen jälkeen ei voida päätellä Aggredin vaikuttavuutta. Näistä syistä seuraavassa tarkastellaan Aggredin asiakkaiden uusimista suhteessa kaltaistettuun verrokkiryhmään. Tarkastelu tehdään ITT-ryhmässä, tutkimusryhmässä, sekä Aggredin onnistuneesti läpikäyneiden ryhmässä. Kusakin tarkastelussa jokaiselle hoitoryhmän henkilölle poimittiin yksi verrokki takaisinpanolla, joka tarkoittaa sitä, että sama verrokki saattoi toimia vertailukohtana useammalle tutkimushenkilölle.

Kun uusien väkivaltarikosten määrää seuranta-aikana verrataan ITT-ryhmässä ja heidän verrokkiryhmässään, tilastollisesti merkittävää eroa rikosten uusimisessa ei havaita millään neljästä vastemuuttujasta. Estimaatit ovat positiivisia, mikä tarkoittaa sitä, että ITT-ryhmä uusii näiden estimaattien mukaan rikoksia hieman verrokkiryhmää useammin, mutta tilastollisesti merkittäviä erot eivät siis ole. Sama pätee myös analyysin, jossa tarkastellaan ainoastaan tutkimusryhmää, eli Aggredin aloittaneita henkilöitä. Näin ollen emme löydä näyttöä siitä, että Aggredi-ohjelma olisi kokonaisuutena arvioiden onnistunut vähentämään väkivaltarikosten uusimista suhteessa kaltaistettuun verrokkiryhmään. Toisaalta on huomionarvoista, että kun analyysi toistetaan ainoastaan Aggredin onnistuneesti läpikäyneessä ryhmässä, ero verrokkiryhmään muuttuu positiivisesta negatiiviseksi. Otoksoon ollessa varsin pieni mikään näistäkin eroista ei saavuta tilastollista merkittävyyttä, mutta ne antavat kuitenkin hieman lisätukea aikaisemman analyysin johtopäätökselle siitä, että onnistuneesti Aggredin suorittaneiden ryhmässä näyttää tapahtuneen myönteistä kehitystä, myös verrattuna vastaavan rikostaustan verrokkeihin.

Taulukko 3 Tutkimus- ja verrokkiryhmän ero rikosten uusimisessa (ATT-estimaatti ja p-arvo)

	Väkivaltarikokset			Kaikki rikokset
	lkm / vuosi	lkm / vapausvuosi	uusineiden osuus	lkm / vuosi
ITT	+0,08	+0,10	+0,05	+0,86
vs. verrokkit (n=141)	(p=0,32)	(p=0,42)	(p=0,34)	(p=0,29)
Tutkimusryhmä	+0,13	+0,07	+0,11	+0,94
vs. verrokkit (n=97)	(p=0,20)	(p=0,69)	(p=0,15)	(p=0,28)
Läpikäyneet	-0,20	-0,47	-0,11	-1,81
vs. verrokkit (n=24)	(p=0,17)	(p=0,11)	(p=0,32)	(p=0,07)

Estimaattia Aggredin toiminnan vaikuttavuudesta heikentää tässä analyysissä todennäköisesti ainakin kaksi asiaa. Ensimmäinen liittyy siihen, että Aggredin onnistuneesti suorittaneiden ryhmä muodostaa verrattain pienen joukon sekä ITT-ryhmästä että tutkimusryhmästä. Jos oletetaan, että tämän ryhmän osalta kyse on todellisesta kausaalista vaikutuksesta (eikä esimerkiksi motivaatioeroista), Aggredin rikollisuutta vähentävän vaikutuksen pitäisi

olla heidän osaltaan varsin merkittävä, jotta se laskisi koko ryhmän rikosuusimisen keskiarvoja tilastollisesti merkittävästi. Toinen tärkeä varaus liittyy verrokkiryhmän uskottavuuteen. Jotta kaltaistusmenetelmällä saataisiin luotettava estimaatti tutkimus- ja vertailuryhmän erosta, täytyisi näiden ryhmien olla keskimäärin samanlaiset ennen hoidon alkua. Vaikka tässä tutkimuksessa käytetty kaltaistusmenetelmä onnistuu tasapainottamaan ryhmät havaittujen kal-

taistutusmuuttujien suhteen, takuuta siitä, että tasapaino keskeisten havaitsemattomien muuttujien suhteen olisi saavutettu, ei ole (Rosenbaum 2010).

Aineisto tarjoaa joitakin mahdollisuuksia tarkastella kaltaistuksen toimivuutta. Muodollinen ehto sille, että kaltaistettujen hoito- ja verrokkiryhmien vertailu voisi tuottaa tietoa ohjelman kausaalista vaikutuksesta uusintarikollisuuteen on se, että niin sanottujen mahdollisten lopputulemien (potential outcomes) tulisi olla riippumattomia hoitoon osallistumisesta (Morgan & Winship 2007). Toisin sanoen hoito- ja verrokkiryhmien rikosten uusimisessa ei pitäisi olla eroa siinä hypoteettisessa tilanteessa, jossa kumpikaan ryhmä

ei saisi hoitoa. Tutkimusryhmän osalta tällaista tietoa ei luonnollisesti voi olla olemassa, eikä oletuksen paikansapitävyyttä sinänsä voi testata. ITT-ryhmässä on kuitenkin mukana huomattava määrä henkilöitä, jotka olisivat voineet olla Aggredin asiakkaita, mutta eivät käytännössä olleet missään kosketuksissa ohjelmaan. Heidän erillinen analyysinsä antaa lisätietoa siitä, onko Aggredin asiakaskunnalle löydettävissä uskottavaa verrokkiryhmää. Mikäli tämän ryhmän rikostaso on seurannan aikana korkeampi kuin verrokeilla, emme voi sanoa sen johtuvan Aggredista, vaan ero liittyy todennäköisemmin kaltaistuksen riittämättömyyteen.

Taulukko 4 Tutkimus- ja verrokkiryhmien ero uusintarikollisuudessa niissä ryhmissä, jotka eivät osallistuneet Aggredi-ohjelmaan (ATT-estimaatti ja p-arvo)

	Väkivaltarikokset			Kaikki rikokset
	lkm / vuosi	lkm / vapausvuosi	uusineiden osuus	lkm / vuosi
ITT-ryhmän hoitamattomat ja korkeintaan kerran käyneet (n=77) vs. verrokkit	+0,23 (p=0,10)	+0,25 (p=0,15)	+0,08 (p=0,26)	+2,11 (p=0,06)
ITT-ryhmän hoitamattomat (n=48) vs. verrokkit	+0,02 (p=0,90)	-0,12 (p=0,56)	+0,04 (p=0,74)	+1,60 (p=0,38)

Taulukossa 4 esitetään tulokset lisäanalyysien pohjalta. Kun tarkastellaan sitä ryhmää, joka ei tullut ollenkaan Aggredin piiriin tai kävi tapaamisessa korkeintaan kerran, havaitaan, että he uusivat väkivaltarikoksensa useammin kuin heidän verrokkinsa (ero ei tosin ole tilastollisesti merkitsevä). Kun analyysistä jätetään pois ne henkilöt, jotka kävivät Aggredissa kerran, estimaatit siirtyvät lähemmäksi nollaa. Näin ollen vaikuttaa siltä, että myös Aggrediin osallistumattomien rikollisuus kasvoi pikemminkin kuin väheni suhteessa verrokkeihin. Jos hyväksytään oletus siitä, ettei Aggredi vaikuttanut tähän ryhmään millään tavalla, antaa analyysi viitteitä siitä, ettei kaltaistus riitä täysin tavoittamaan todellisia eroja Aggredin potentiaalisen asiakaskunnan ja heidän verrokkiansa rikosalttiudessa ennen seurannan alkua.

5 Pohdintaa

Tutkimuksen rajoitukset. Tutkimuksella, jossa hoidon vaikuttavuutta arvioidaan havainnoivalla (vs. kokeellisella) aineistolla on useita rajoitteita. Kaikkein rikollisuutta koskevaan rekisteritutkimukseen liittyvä ilmeinen rajoite on se, että vain viranomaisten tietoon tullutta rikollisuutta voidaan tutkia. Toinen rajoite on se, että aineisto ei sisältänyt tietoa tutkimushenkilöiden ja verrokeiden kuolemista tai muutosta ulkomaille, joten joiltain osin seuranta-aika, johon rikollisuuden määrää suhteutetaan, voi olla liian pitkä. Tärkein rajoite liittyy kuitenkin vertailuryhmän muodostamiseen ja vertailun mielekkyyteen. Kaltais-

tusmenetelmät voivat luonnollisesti käyttää apunaan ainoastaan havaittuja/mitattuja muuttujia, kun todellisuudessa hoitoon valikoitumiseen ja vastetapah-tumaan tyypillisesti vaikuttaa suuri joukko tekijöitä, joita ei joko 1) ole mitattu tai 2) tiedostettu ylipäätään (Rosenbaum 2010). Kun ohjelmaan osallistuminen on vapaaehtoista, halukkuudella muutokseen on väistämättä vaikutus siihen, ketkä päätyvät ohjelman pariin ja ketkä eivät.

Tässä tutkimuksessa Aggredin tutkimusjoukon rikollisuutta verrattiin iän, sukupuolen ja rikostaustan perusteella kaltaistettuun verrokkiryhmään. Vaikka kaltaistus onnistuikin näiden muuttujien suhteen, on edelleen mahdollista, että havaitsemattomat erot tutkimus- ja verrokkiryhmän välillä selittävät ryhmien välistä eroa rikosten uusimisessa. Aggredin toiminnan kohdistumisen huomioon ottaen nämä erot saattavat liittyä mm. maantieteellisiin eroihin rikollisuudessa, sillä on mahdollista, että pääkaupunkiseudun rikollisuus eroaa rakenteeltaan muusta maasta merkittävästikin esimerkiksi suhteessa järjestäytyneeseen rikollisuuteen. Tällöin on mahdollista, että Aggredin asiakkaat olivat jo lähtötilanteessa väkivalta-alttiimpia kuin vastaavalla rikostaustalla varustetut verrokkit muualla Suomessa. Mikäli näin on, arvio Aggredi-ohjelman tehosta on konservatiivinen. Toisaalta on kuitenkin todettava, ettei parempi kaltaistus välttämättä johtaisi siihen loppupäätelmään, että Aggredilla on rikollisuutta vähentävä vaikutus. Käytettävissä olevien tietojen perusteella vaikuttaa kuitenkin uskottavalta, että verrokkiryhmän väkivalta-alttius oli todennäköisemmin Aggredin asiakkaita matalampi kuin korkeampi lähtötilanteessa.

Yhteenveto. Rikoksantorjuntaohjelmien vaikuttavuutta on arvioitu Suomessa toistaiseksi varsin harvoin tilastollisin menetelmin. Aggredin kiistämättömäksi ansioksi on laskettava heidän halukkuutensa altistaa työmuotonsa vaikuttavuusarvioinnin kohteeksi. Myönteistä on se, että Aggredi-ohjelma tavoittaa hyvin kohdejoukkonsa, sillä Aggredin asiakkailta on analyysin perusteella varsin raskas rikostausta. Kyseessä on siis monin tavoin haastava kohderyhmä.

Ohjelman tulokset ovat hyviä sen läpikäyneessä ryhmässä sekä väkivaltarikollisuuden että muunkin rikollisuuden vähenemisen muodossa. Eron läpikäyneiden ja ohjelman keskeyttäneiden välillä liittyy toisaalta myös ohjelman keskeisin haaste: suuri osa ohjelman aloittaneista ei suorita sitä onnistuneesti loppuun, ja tämä ryhmä uusii väkivaltarikoksensa usein. Samainen ryhmä nostaa koko tutkimusryhmän rikosuusimisen siinä määrin korkealle tasolle suhteessa verrokkiryhmään, ettei arvioinnin perusteella voida todeta ohjelmalla olevan kokonaisuutena arvioiden väkivaltarikollisuutta vähentävää vaikutusta. Ilman tällaista tulosta emme voi varmuudella rajata pois sitä mahdollisuutta, että ohjelman onnistuneesti suorittaminen heijastaisi havaitsemattomia eroja esimerkiksi motivaatiossa ja pyrkimyksessä kohti rikoksetonta elämää. Toisaalta lisäanalyysit antoivat viitteitä siitä, että verrokkiryhmä ei todennäköisesti ollut lähtökohdiltaan keskimäärin yhtä väkivalta-altista kuin tutkimusryhmä. Näitä vastakkaisia näkökohtia punniten arvioinnin loppupäätelmä on se, että tulokset ovat ohjelman läpikäyneiden osalta siinä määrin lupaavia, että Aggredi-ohjelmaa kannattaa jatkaa ja kehittää edelleen.

Tämän arvioinnin perusteella ei voida sanoa, että Aggredi olisi sellainen väkivaltatyön muoto, jonka laajentamisella esimerkiksi muille paikkakunnille saataisiin varmasti hyviä tuloksia. Tällainen päätelmä vaatisi mielellään ITT-asetelmaan perustuvaa näyttöä ohjelman rikollisuutta vähentävästä vaikutuksesta. Eräs tärkeä näkökohta on se, että Aggredin asiakaskunta aiheuttaa rikoskäyttäytymisellään hyvin merkittäviä yhteiskunnallisia kustannuksia. Jos Aggredin ansiosta vältettäisiin esimerkiksi yksi vankeusvuosi, tarkoittaisi se noin 65 000 euron kustannussäästöä (Hinkkanen 2012). Vaikka mielekkään kustannus-hyötyanalyysin esittäminen vaatisi luotettavampaa estimaattia ohjelman tehosta, voidaan kuitenkin todeta, että onnistuessaan Aggredi-ohjelmalla voi olla hyvin merkittäviäkin kustannusvaikutuksia.

Mikäli vastaaviin ohjelmiin kohdistuvia arviointeja halutaan jatkossa toteuttaa, olisi tämän tutkimusprosessin kokemusten valossa pyrittävä siihen, että tutkimus- ja vertailuryhmä satunnaistetaan ulkopuolisen tahon toimesta ennen hoidon aloittamista. Yksi vaihtoehto olisi verrata kahta eri interventiota. Tällainen menettely voisi olla yleisesti hyväksyttävämpi kuin vaihtoehto, jossa

verrokkiryhmä jätetään tyystin ilman hoitoa. Mikäli tällainen asetelma toteutetaan Aggredin tapauksessa, olisi lisäksi pohdittava sitä, mikä on se ryhmä, jolle ohjelmaa lähdetään lähtökohtaisesti tarjoamaan. Olisiko esimerkiksi järkevää tarjota ohjelmaa ainoastaan ehdottomasta vankeudesta vapautuville, jotka kiinnittyivät tämän analyysin perusteella ohjelmaan muita paremmin?² Jos ohjelman kohdejoukossa on merkittävä määrä henkilöitä, jotka eivät motivoitu aloittamaan ohjelmaa, olisi myös satunnaistetussa asetelmassa käytännössä vaikea havaita tilastollisesti merkitseviä eroja tutkimus- ja verrokkiryhmien välillä pienessä aineistossa, vaikka ohjelmalla olisikin todellisuudessa väkivaltarikollisuutta vähentävä vaikutus.

Mikäli satunnaistaminen ei ole mahdollista, tulisi vaihtoehtoisesti pyrkiä keräämään enemmän taustatietoa sekä tutkimus- että vertailuryhmistä, ja hyödyntää tätä tietoa kaltaistamisessa. Tällöin on kuitenkin muistettava, ettei hyväänkään havainnoivaan aineistoon perustuva kaltaistus korvaa satunnaistamista. Käytännössä tällaisen lisäinformaation saaminen edellyttäisi yhteistyötä eri viranomaistahojen kesken, ja vähintään tarkempaa tietoa ohjaavien tahojen toiminnasta ja siitä, millä perusteilla henkilöitä lähetetään Aggrediin. Tällöin tulisi varmistaa, että kerättävä lisäinformaatio on riittävän laadukasta ja yhdenmukaista, ja kertoo henkilöiden väkivalta-alttiudesta jotain sellaista, joka ei ole tarttunut rikostaustaa kuvaaviin mittareihin. Käytännössä tällaisen tiedon systemaattinen kerääminen voi olla vaikeaa.

Yleisemmällä tasolla voi todeta, että tässä tutkimuksessa toteutetun kaltainen asetelma voisi puutteistaan huolimatta olla jatkossa käyttökelpoinen rikoksantorjuntaohjelmien arvioinneissa, varsinkin silloin kun satunnaistaminen ei ole lainsäädäntöön tai ohjelman toteutukseen liittyvistä syistä mahdollista. Perinteiseen regressioanalyysiin verrattuna kaltaistusmenetelmät ovat hyviä erityisesti siksi, että niissä on mahdollista eksplisiittisesti erottaa tutkimus- ja verrokkiryhmien kaltaistaminen varsinaisesta ohjelman tehon analyysistä. Havainnoivia aineistoja hyödyntäviin menetelmiin piilee aina riski siitä, että kaltaistuksen onnistumista arvioidaan analyysin lopputuloksen, ei tasapainotuksen onnistumisen perusteella. Erityisesti ohjelmat, joissa kohderyhmä muistuttaa rikostaustaltaan keskimääräistä rikoksentekeijää, ja sopivia verrokkeja löytyy paljon, ovat tämän kokemuksen valossa potentiaalisia kaltaistetun arvioinnin kohteita. On lisäksi näyttöä siitä, että ohjelmat, joissa hoitoon ohjautuminen ei perustu henkilön omaan vapaaehtoisuuteen, vaan objektiivisiin ja tutkijan tuntemiin kriteereihin, saattavat sopia paremmin arvioitavaksi kaltaistamismenetelmien avulla (Hämäläinen, Uusitalo & Vuori 2008). Sitä, miten valikoitumisharha sillä hetkellä tehtävässä arvioinnissa vaikuttaa, täytyy kuitenkin aina pohtia tapauskohtaisesti.

² 31 prosenttia vankiloista tulleista suoritti ohjelman loppuun, kun vastaava osuus yhdyskuntaseuraamustoimistojen kautta tulleista oli 23 prosenttia ja muualta Aggrediin ohjatuista 15 prosenttia.

Lähteet

- Abadie, Alberto & Leber Herr, Jane & Imbens, Guido W. & Drukker, David M. 2004. NNMATCH: Stata module to compute nearest-neighbor bias-corrected estimators. <http://ideas.repec.org/c/boc/bocode/s439701.html>.
- Austin, Peter C. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research* 46: 399–424.
- Becker, Sascha O. & Ichino, Andrea. 2002. Estimation of average treatment effects based on propensity scores. *Stata Journal* 2(4): 358–377.
- Caliendo, Marco & Kopeinig, Sabine. 2005. Some practical guidance for the implementation of propensity score matching. IZA Discussion Paper No. 1588. Bonn: Institute for the Study of Labor.
- Diamond, Alexis & Sekhon, Jasjeet S. 2013. Genetic matching for estimating causal effects: a general multivariate matching method for achieving balance in observational studies. *The Review of Economics and Statistics* 95(3): 932–945.
- Gupta, Sandeep K. 2011. Intention-to-treat concept: A review. *Perspectives in Clinical Research* 2(3): 109–112.
- Hinkkanen, Ville. 2012. Rikollisuuden kustannukset. Teoksessa *Rikollisuustilanne 2011. Rikollisuus ja seuraamusjärjestelmä tilastojen valossa*. Helsinki: Oikeuspoliittinen tutkimuslaitos.
- Hämäläinen, Kari & Uusitalo, Roope & Vuori, Jukka. 2008. Varying biases in the matching estimates: evidence from two randomized job search training experiments. Helsinki: Valtion taloudellinen tutkimuskeskus.
- Kekki, Tuula & Salakka, Petri. 2012. Aggredi – Väkiältä ei ole mulle mikään ongelma. Helsinki: Aggredi.
- Konttila, Arja & Tyni, Sasu. 2011. OMA-ohjelman suorittaneiden väkivaltarikollisten uusintarikollisuus sekä ohjelman välittömät vaikutukset. Helsinki: Rikosseuraamuslaitos.
- Laine, Markus & Tuominen, Martti. 2012. Ne ymmärtää katujen todellisuutta – Kokemuksia toisenlaisesta väkivaltatyöstä. *Kvartti* 1/2012. Helsinki: Helsingin kaupungin tietokeskus.
- Leuven, Edwin & Sianesi, Barbara. 2003. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. <http://ideas.repec.org/c/boc/bocode/s432001.html>.
- Morgan, Stephen L. & Winship, Christopher. 2007. Counterfactuals and causal inference. *Methods and principles for social research*. Cambridge: Cambridge University Press.
- Rosenbaum, Paul R. 2010. *Design of observational studies*. New York: Springer.
- Sarna, Seppo. 2012. Sanasto, klinisen biostatistiikan kurssi.
- Sekhon, Jasjeet S. 2011. Multivariate and propensity score matching software with automated balance optimization: the Matching package for R. *Journal of Statistical Software* 42(7): 1–52.
- Sirén, Reino. 2012. Ehdoton vankeus, yhdyskuntapalvelu ja rikosuusiminen. Ehdoton vankeus ja yhdyskuntapalvelu törkeän rattijuopumuksen seuraamuksina. Helsinki: Oikeuspoliittinen tutkimuslaitos.
- Skardhamar, Torbjørn & Savolainen, Jukka. 2014. Changes in criminal offending around the time of job entry: a study of employment and desistance. *Criminology*, advance access.