



Vankeusaikaisten kuntoutusohjelmien vaikutusarvioinnit ja arvioitavuus

SASU TYNI & MIKKO AALTONEN

Suomalaisissa vankiloissa on tehty systemaattista ohjelmatoimintaa noin 20 vuoden ajan. Ohjelmatoiminnan tärkeimpänä tavoitteena on edistää rikoksetonta elämäntapaa, ja siten vähentää vapautumisen jälkeistä uusintarikollisuutta. Uusintarikollisuusvaikutusten todentaminen edellyttää vaikutustutkimuksiin perustuvaa tutkimusnäyttöä, jota on vähitellen alkanut kertymään myös Suomesta. Tässä artikkelissa käydään läpi yksittäisten ohjelmien uusintarikollisuusvaikutuksia koskevien kotimaisten arviointitutkimusten keskeiset tulokset, sekä pohditaan ohjelmatoiminnan vaikutusten arvioinnin mahdollisuuksia tilastollisten tutkimusmenetelmien, tutkimusaineistojen ja ohjelmatoiminnan volyymin näkökulmista. Ohjelmatoiminnan vaikutuksista uusintarikollisuuteen Suomessa ei ole toistaiseksi saatu selvää näyttöä. Vaikka ohjelmien vaikutusten tutkimiseen tarvittavat aineistot ovatkin parantuneet 2000-luvulla selvästi, uskottavaa arviointia vaikeuttavat edelleen hyvien tutkimusasetelmien puute sekä yksittäisten ohjelmien pienet osallistujamäärät. Vaikutusarviointien menetelmällisessä laadussa on parantamisen varaa. Koska yksittäisen ohjelman vaikutus uusintarikollisuuteen on saatavilla olevan tutkimustiedon valossa luultavasti aika maltillinen, kovin pienillä tutkimusaineistoilla vaikutuksia ei kannata jatkossa yrittää tilastollisesti tutkia. Pienet vaikutukset eivät kuitenkaan tarkoita sitä, että ohjelmat eivät voisi olla kannattavia ja kustannustehokkaita. Merkittävä osa nykyisestä ohjelmatoiminnasta on osallistujamääriltään niin pienimuotoista, ettei niiden toimivuudesta tai toimimattomuudesta ole helppo saada luotettavaa tietoa. Tämä tilanne haastaa tutkimustietoon perustuvan ohjelmatoiminnan kehittämistä.

English summary at the end of the article

Avainsanat: vankilakuntoutus – uusintarikollisuus – vaikuttavuuden arviointi – tilastollinen voima – kausaalipäätely

JOHDANTO

Anglo-amerikkalaiseen vankeinhoitoon juurtui 1970-luvulla käsitys siitä, että vankilassa järjestettävällä toiminnalla ei voida vähentää vankien uusintarikollisuutta ("Nothing Works!", Martinson, 1974; Martinson, 1979). 1980-luvun lopusta alkaen vankeinhoidon mahdollisuuksiin suhtauduttiin jo optimistisemmin. Meta-analyysien perusteella etenkin kognitiivis-behavioraalisiin lähtökohtiin perustuvien ohjelmien havaittiin vähentävän vankien uusintarikollisuutta (Andrews ym., 1990). Tutkimuksissa siirryttiin vertailemaan erilaisten ohjelmien ja interventioiden uusimisvaikutuksia ("What Works?"). Vankeinhoidon lippulaivoiksi nousivatkin kuntoutusohjelmat: pitkien vankilatuomioiden havaittiin jopa lisäävän vankien uusimisriskiä, mutta tietyillä ohjelmilla todettiin olevan uusimisriskiä vähentävä vaikutus (Andrews ym., 1990).

Pohjoismaissa ja Suomessa rikosseuramusalalla mullistaviin tuloksiin suhtauduttiin maltillisemmin. Ohjelmia otettiin käyttöön, mutta vankeinhoidon kuntoutus sisälsi myös muita perinteisempiä toimintamuotoja kuten päihdekuntoutusta, vankityötä, vankien koulutusta sekä muuta sosiaalista kuntoutusta kuten arkielämän perustaitojen harjoittamista (Tyni, 2015). Vankiloiden ohjelmatoiminnan alkaminen voidaan ajoittaa Suomessa 1990-luvun loppuun, jolloin käynnistettiin ensimmäiset kuntoutusohjelmat. Kuntouttavien ohjelmien perusteet rantautuivat Suomeen pääasiassa Kanadasta ja Isosta-Britanniasta (Lavikkala, 2011). Merkittäviksi ajankohdiksi voidaan nostaa vankien päihteiden käytön vähentämiseen tähtäävän toiminnan lisäämiseksi ja kehittämiseksi asetetun projektin (Koski-Jännes, 1995) toimenpide-ehdotukset ja jatkokehityshankkeet sekä Vankeinhoidon

koulutuskeskuksessa vuonna 2001 pidetty seminaari, jossa sitouduttiin uuden suuntautumistavan lähtökohtiin ja ajatteluun (Tyni, 2015; Lavikkala, 2011).

Vankiloiden tämänhetkinen ohjelmatyö voidaan jakaa erilaisiin motivointiohjelmiin, rikosperusteisiin ohjelmiin, päihdeohjelmiin ja yleisohjelmiin. Ohjelmat toteutetaan joko yksilö- tai ryhmätyönä tai näiden yhdistelmänä. Ohjelmatyössä ei ole kyse varsinaisesta terapiasta, vaan strukturoidusta työskentelymuodosta, joilla pyritään vähentämään vankien uusintarikollisuutta vankilasta vapautumisen jälkeen. Suurin osa ohjelmatoiminnasta perustuu kognitiivis-behavioraaliseen lähestymistapaan (Knuuti & Vogt-Airaksinen, 2013). Viimeisimpinä vuosina vankeinhoito on sisältänyt myös ns. Hyvän elämän malliin (Good Lives Model) eli voimavarasuuntautuneeseen viitekehykseen perustuvia ohjelmia (Tyni, 2015).

Vankiloiden rikosperusteisten ohjelmien ja yleisohjelmien ensisijaisena tarkoituksena on parantaa vankien valmiuksia rikoksettomaan elämäään. Ohjelmien keskeisimpänä vaikutusvasteena voidaan täten pitää vapautumisen jälkeistä uusintarikollisuutta. Ohjelmatyön yleistyttyä onkin perusteltua kysyä, onko Suomessa toteutetuilla ohjelmilla ollut vaikutusta uusintarikollisuuteen? Vuosina 2000–2012 vapautuneista suomalaisista vangeista keskimäärin 79 prosenttia syyllistyi uusiin rikoksiin viiden vuoden sisällä vapautumisesta, ja noin 58 prosenttia vangeista päätyi takaisin vankilaan uusista rikoksista (Tyni, 2020). Myönteistä on se, että uusintarikollisuusaste on molemmilla edellä mainituilla mittareilla laskenut jonkin verran. Tästä löydöksestä ei kuitenkaan suoraan voida päätellä ohjelmatoiminnan mahdollista vaikutusta havaittuun muutokseen. Yksittäisten ohjelmien uusintarikollisuusvaikutusten selvittäminen

vaatii vuositilastoja ja indikaattoreita syvem-
mälle menevän vaikutustutkimuksen teke-
mistä.

Tämän artikkelin tarkoituksena on muodos-
taa kokonaiskäsitely jo tehdyistä ohjelmatoi-
minnan arvioinneista, sekä laajemmin käsi-
tellä ohjelmatoiminnan vaikutuksia koskevan
tutkimuksen realistisia mahdollisuuksia ja
reunaehdoja. Keskitymme erityisesti siihen,
miten vaikutuksia pitäisi tilastollisesti tutkia,
ja siihen, miksi ohjelmien uusintarikollisuus-
vaikutusten luotettava arviointi on käytän-
nössä vaikeaa. Aloitamme artikkelin kuvaam-
alla kuntoutusohjelmien määrää ja laatua
suomalaisissa vankiloissa 2010-luvulla. Tutki-
musaineistona käytämme Rikosseuraamus-
laitokselta saatuja ohjelmatilastoja, joihin on
yhdistetty vankien ajankäyttöön liittyviä tun-
nuslukuja. Näiden tietojen esittämisen tarkoi-
tuksena on ensisijaisesti osoittaa, millaisia ra-
joituksia ohjelmatoiminnan nykyinen vo-
lyymi asettaa tilastolliselle vaikutustutki-
mukselle, ja samalla tarjota vaikutustutki-
musta suunnitteleville tutkijoille aikaisem-
min julkaisemattomia tietoja ohjelmatoimin-
nan määrästä ja laadusta. Seuraavaksi ku-
vaamme Suomessa tähän saakka tehtyjen oh-
jelmien vaikutuksia koskevien arviointitutki-
musten päätulokset. Tässä tutkimuksen
osassa aineistona ovat kyseisten arviointien
julkaistut raportit.

Tämän jälkeen keskustelemme esimerkkien
kautta ohjelmatyön vaikuttavuuden arvioin-
nin yleisistä edellytyksistä sekä keskeisistä
ongelmista tutkimusasetelmien, saatavilla
olevien aineistojen ja tilastollisen voiman nä-
kökulmasta. Vaikuttavuuden kriteeriksi nos-
tamme artikkelissa ohjelmien uusintarikolli-
suusvaikutukset (ks. myös Aaltonen & Tyni,
2017), mutta artikkelissa kuvaaviamme vai-
kutustutkimuksen peruslähtökohtia ja ope-
tuksia voidaan hyödyntää monilta osin myös

muissa vankeinhoidon toimintaa tai laajem-
min erilaisten toimenpiteiden vaikutuksia
selvittävässä kriminologisissa tutkimuksissa,
vaikuttavuuden kriteeristä ja tarkasteltavasta
vasteesta riippumatta.

OHJELMATYÖN MÄÄRÄ JA LAATU

Rikosseuraamuslaitoksen Vankitietojärjes-
telmä (VATI) sisältää kattavasti tutkinta-,
sakko- ja vankeusvankien sekä valvontaran-
gaistukseen ja nuorisorangaistukseen tuomit-
tujen henkilöiden seuraamusten suorittami-
seen liittyviä tietoja. Vankiloiden ohjelma-
työn määrää ja laatua voidaan seurata Vanki-
tietojärjestelmässä sekä vankien ajankäyttö-
tietojen että ohjaajien suoritekirjausten pe-
rusteella. Vankien ajankäytön seuranta pe-
rustuu vankipalkkausjärjestelmään. Vankien
ansiokorteista voidaan eritellä tuntiperusteis-
esti aktiivitoimintojen osuus sellaisesta
ajankäytöstä, jolloin vanki on poissa aktiivi-
toiminnoista esimerkiksi sairastumisen tai
käräjämatkojen vuoksi.

Rikosseuraamuslaitoksen ajankäyttötilas-
toinnin mukaan aktiivitoimintojen osuus on
noussut hieman 2010-luvun aikana. Vuonna
2010 vangeista keskimäärin 49 prosenttia
osallistui arkipäivisin vankitoimintoihin;
vuonna 2019 toimintoihin osallistui keski-
määrin 55 prosenttia vangeista. Työtoimin-
nan osuus oli noin kolmannes (32 %) kirja-
tusta kokonaisajankäytöstä. Vankikoulutuk-
sen osuus oli noin 8 prosenttia ja muiden ak-
tiivitoimintojen noin 15 prosenttia kokonais-
ajankäytöstä. (Rikosseuraamuslaitos A.)
Näistä tilastoista voidaan päätellä, että kai-
kista aktiivitoiminnoista ainoastaan pieni osa
on ohjelmatoimintaa, koska ohjelmatoiminta
sisältyy luokittelussa muihin aktiivitoimin-
toihin.

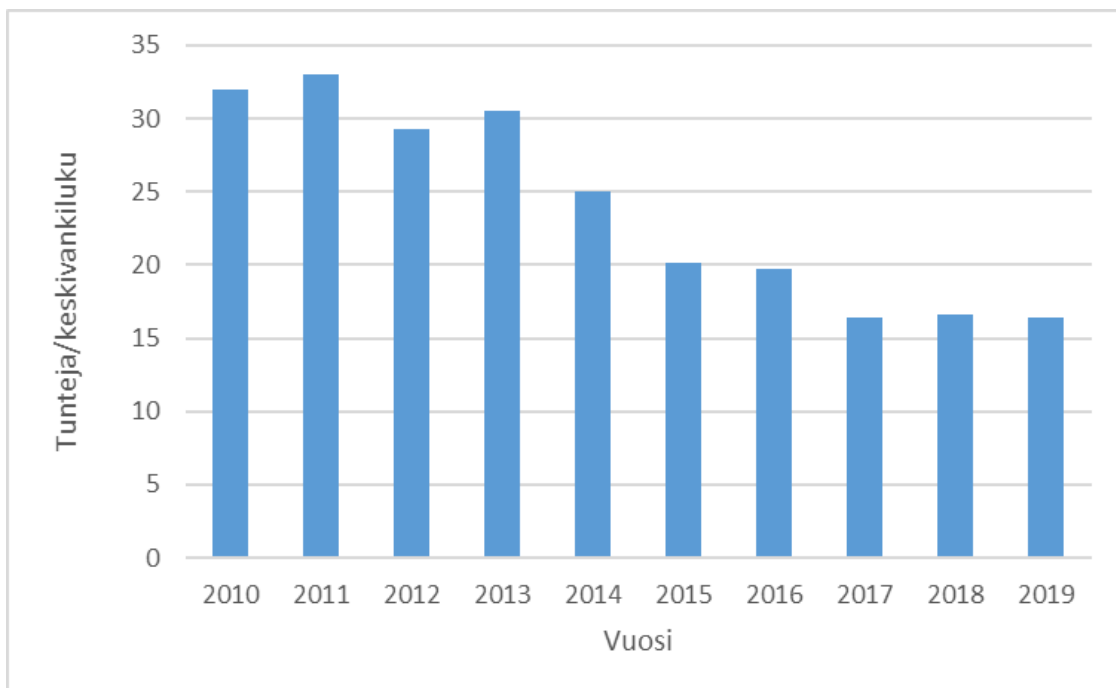
Rikosseuraamuslaitoksen tilastointia voidaan laajentaa siten, että kullekin yksittäiselle vangille lasketaan aktiivian osuus koko vankilakauden kokonaisajankäytöstä (vrt. Tyni & Blomster, 2012). Artikkeliamme varten muodostimme kaikista vuosina 2010–2019 vapautuneista vankeusvangeista perusaineiston (N = 27 726), johon liitimme Vankitietojärjestelmästä vankien ansiokorttitiedot sekä kuntoutusohjelmien suorittamiseen liittyneet aineistot. Kuvaamme seuraavassa ohjelmiin kirjattua ajankäyttöä ja ohjelmien suoritelmääriä.

2010-luvulla vankien aktiivisen ajankäytön osuus on noussut lähes kymmenen prosenttiyksikköä: vuonna 2010 vapautuneilla vankeusvangeilla keskimäärin 29 prosenttia vankeusajasta oli aktiiviaikaa, vuonna 2019 osuus oli noussut 38 prosenttiin. Vankien kokonaisajankäytöstä on kuitenkin hankala erottaa kuntoutusohjelmiin kirjattua aktiiviaikaa, koska ohjelmien toimintoluokitteluja ja kirjaamisohjeistuksia on muutettu tarkastelujaksolla useampaan kertaan. Esimerkiksi

eräitä yksittäisiä ohjelmia ei ollut mahdollista erottaa muista ohjelmista 1.10.2010 tapahtuneen Vankitietojärjestelmän päivityksen jälkeen. Vuoden 2016 tammikuusta alkaen ohjelmien luokittelu typistettiin lopulta siten, että akkreditoitujen eli hyväksymismenettelyyn läpikäyneet päihdeohjelmat, uusintarikollisuuteen vaikuttavat ohjelmat sekä motiiviohjelmien yhdistettiin yhdeksi toimintoluokaksi.

Huomionarvoista on, että akkreditoitujen ohjelmien suoritelmäärät ja näihin kohdistetut ajankäyttötunnit ovat alentuneet selvästi viimeisten vuosien aikana. Vuonna 2010 akkreditoituihin ohjelmiin kirjattiin yhteensä noin 105 000 suoritettua tuntia, kun vuonna 2019 vastaava tuntimäärä oli pudonnut yli puolella eli noin 48 000 tuntiin. Keskimääräiseen vankilukuun suhteutettuna akkreditoitujen ohjelmien tilanne näyttääkin varsin huolestuttavalta (Kuvio 1).

Ajankäyttötietojen lisäksi kuntoutusohjelmien määriä voidaan tarkastella myös ohjaa-

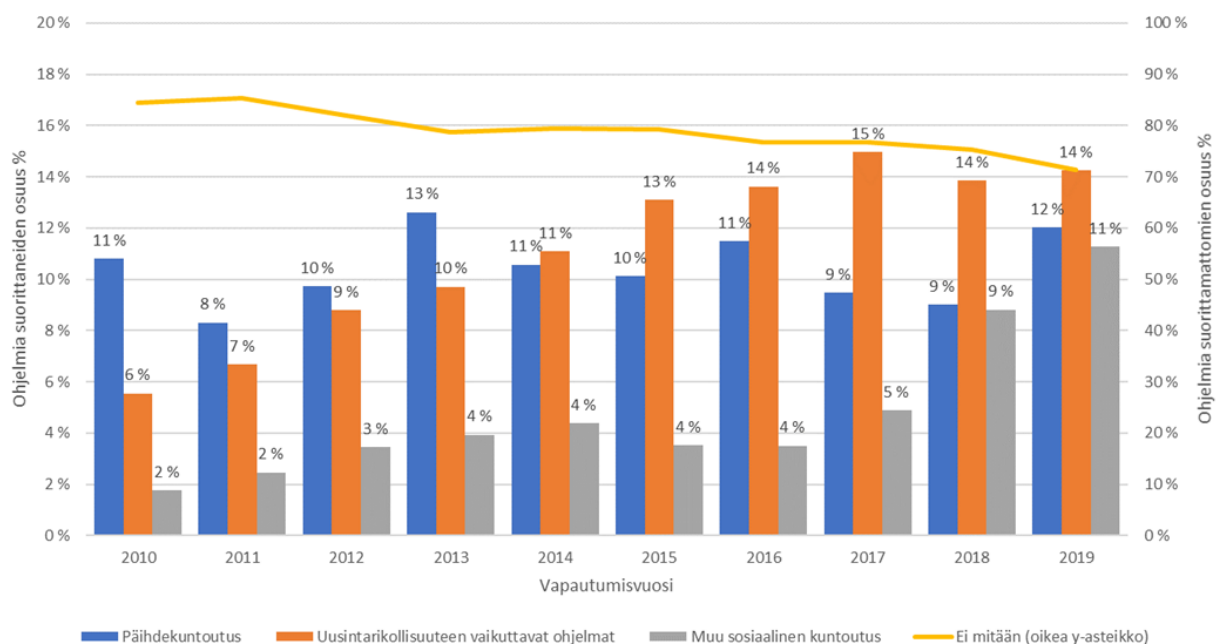


Kuvio 1. Akkreditoituihin ohjelmiin kohdistettu ajankäyttö suhteutettuna keskivankilukuun vuosina 2000–2019 (lähde: Rikosseuraamuslaitos).

jien tekemien suoritekirjausten avulla. Kuntoutusohjelmat on luokiteltu Vankitietojärjestelmän ohjelmaosuudessa päihdekuntoutusohjelmiin, uusintarikollisuuteen vaikuttaviin ohjelmiin sekä sosiaaliseen kuntoutukseen. Päihdeohjelman on suorittanut 2010-luvulla noin 350–550 vankia vuodessa. Uusintarikollisuuteen vaikuttavien ohjelmien suorittemäärät ovat vaihdelleet noin 400–600 vangin välillä. Sosiaalisen kuntoutuksen vuosittainen määrä on sen sijaan noussut tasaisesti koko 2010-luvun ajan; vuonna 2010 sosiaaliselle kuntoutukselle kirjattiin hieman yli 100 suorituskertaa, mutta vuonna 2019 suorituskerta oli jo lähes 550.

Päihdekuntoutusohjelmien ja uusintarikollisuuteen vaikuttavien ohjelmien suorittemäärissä on ollut suurta vuosittaista vaihtelua. Joidenkin ohjelmien tai vankiryhmien osalta tilanne on kuitenkin melko hyvä: esimerkiksi

seksuaalirikollisille suunnattu kognitiivis-behavioraalinen STOP-ohjelma poistettiin vankiloista vuonna 2018 ja korvattiin STEP-ohjelmalla¹, mutta seksuaalirikollisten kuntoutusmäärät ovat pysyneet maltillisen tasaisina. STEP-ohjelman lisäksi seksuaalirikosvangit ovat osallistuneet vankilassa Uusi Suunta-yksilöohjelmaan. Viisi Keskustelua Muutoksesta -motivointiohjelman (VKM) suorittemäärät ovat sen sijaan jopa nousseet merkittävästi². Kyseisen ohjelman suorittemäärien nousu selittää suurelta osin sitä, miksi uusintarikollisuuteen vaikuttavien ohjelmien määrä nousi 2010-luvun ensimmäisten vuosien aikana. Silmiinpistävintä oli se, että etenkin pitkäkestoisten ohjelmien suorittemäärät olivat pudonneet huomattavasti vuosikymmenen ensimmäisiin vuosiin verrattuna.



Kuvio 2. Vuosina 2010-2019 vapautuneiden, eri ohjelmia suorittaneiden vankien lukumäärä (lähde: Rikosseuraamuslaitos)

¹ STEP-ohjelma toteutetaan riski-, tarve- ja vastaavuusperiaatteen lisäksi motivoivan haastattelun, positiivisen psykologian ja Hyvän elämän mallin viitekehyksessä.

² Viisi Keskustelua Muutoksesta -ohjelma perustuu motivoivan haastattelun menetelmään.

Jos ohjelmien suorittamisen yleisyyttä tarkastellaan yksittäisen vapautuneen vangin tasolla (Kuvio 2), havaitaan, että ohjelmia suorittamattomien vankien osuus on 2010-luvulla laskenut 85 prosentista 71 prosenttiin (Kuvio 2). Edelleenkin enemmistö vangeista jää kuntoutusohjelmien ulkopuolelle, mutta kokonaisuutena arvioiden tilanne vaikuttaa parantuneen hieman. Päihdekuntoutuksessa olleiden osuus on pysynyt kohtalaisen vakaana, uusintarikollisuuteen vaikuttavissa ohjelmissa tai muussa sosiaalisessa kuntoutuksessa olleiden puolestaan nousut.

Taulukossa 1 on vielä esitetty eräitä vaikuttavuuden arvioinnin kannalta keskeisiä taustatietoja eri ohjelmatyypejä suorittaneista ja kokonaan ohjelmiin osallistumattomista vapautuneista vangeista samalla seurantajak-solla. Keskeisin ero näiden kahden ryhmän välillä on se, että ohjelmiin osallistuneet van-

git suorittavat selvästi osallistumattomia vankeja pidempiä tuomioita. Tämä johtunee siitä syystä, että lyhyiden tuomioiden aikana ohjelmia ei yksinkertaisesti ehditä suorittaa. Yllättävää ei ole myöskään se, että eri ohjelmatyyppeihin osallistuneiden vankien aktiivijajan osuus on osallistumattomia vankeja suurempi. Kiinnostavaa sen sijaan on, että huomattavasti pidemmistä tuomioista huolimatta ohjelmiin osallistuneiden vankien vankeuskertaisuus on samalla tasolla osallistumattomien vankien kanssa. Eri ohjelmatyyppeihin osallistuneiden vankien taustoissa on jopa yllättävän vähän eroja. Tässä tarkastelussa on tosin huomioitava osittainen päällekkäisyys ryhmien välillä, sillä sama vanki on voinut suorittaa vankeuskaudellaan useamman ohjelman.

Taulukko 1. Eri ohjelmia suorittaneiden, vuosina 2010–2019 vapautuneiden vankien taustatiedot ja lukumäärät.

	Vapautumisvuosi										Yht.
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	
Päihdekuntoutus											
Kertaisuus (ka)	4,5	5,2	4,5	4,4	4,5	4,6	4,7	5,2	4,5	4,9	4,7
Ikä vapautuessa (ka)	35,2	36,6	36,1	35,8	36,4	36,7	37,2	37,4	35,0	37,0	36,3
Tuomion pituus (md)	14	16	15	13	15	15	13	18	16	9	14
Aktiivijajan osuus (%)	41	42	42	39	40	41	41	40	42	41	41
Vankien määrä	454	331	363	452	358	340	370	303	285	377	3633
Uusintarikollisuuteen vaikuttavat ohjelmat											
Kertaisuus (ka)	4,2	5,0	4,2	3,9	4,0	4,2	3,9	4,0	3,8	3,4	4,0
Ikä vapautuessa (ka)	34,2	36,9	35,0	35,1	35,7	36,2	36,0	36,4	36,0	36,0	35,8
Tuomion pituus (md)	18	18,5	18	18	18	17	18	19	18	19	18
Aktiivijajan osuus (%)	40	41	44	43	43	43	44	43	47	49	44
Vankien määrä	233	266	328	348	376	440	439	479	439	447	3795
Muu sosiaalinen kuntoutus											
Kertaisuus (ka)	5,0	3,2	4,3	3,8	4,7	4,1	5,2	4,2	4,4	4,9	4,4
Ikä vapautuessa (ka)	39,7	37,1	36,6	35,3	36,9	36,7	38,5	36,4	36,2	37,4	36,9
Tuomion pituus (md)	18	20	17	15	17	20	22	20	12	12	16
Aktiivijajan osuus (%)	38	40	45	43	41	42	41	45	48	48	44
Vankien määrä	74	98	129	141	148	118	113	156	279	354	1610
Ei mitään edellisistä											
Kertaisuus (ka)	4,3	4,5	4,4	4,5	4,5	4,6	4,6	4,8	4,6	4,8	4,5
Ikä vapautuessa (ka)	36,8	36,5	36,6	37,1	37,2	37,4	37,4	37,3	37,4	37,4	37,1
Tuomion pituus (md)	2	2	2	3	2	2	2	3	2	2	2
Aktiivijajan osuus (%)	26	26	28	28	26	28	30	32	32	34	29
Vankien määrä	3551	3403	3053	2818	2690	2658	2475	2453	2384	2241	27726

SUOMALAISET ARVIOINTITUTKIMUKSET

Suomessa on toistaiseksi arvioitu tilastollisin menetelmin viidessä eri tutkimuksessa kolmea eri ohjelmaa, joiden ensisijainen tarkoitus on vähentää vankien uusintarikollisuutta. Arvioidut ohjelmat ovat Cognitive Skills (CS), Omaehtoisen muutoksen ohjelma (OMA) ja STOP-ohjelma. Kuvaamme seuraavassa näiden vaikutustutkimusten tutkimusasetelmat ja keskeiset tulokset. Esittelemme myös lyhyesti vastaavien kansainvälisten ohjelmien vaikuttavuutta tutkineet lähteet.

Cognitive Skills

Ensimmäisessä suomalaisessa ohjelmatoimintaa koskevassa vaikutustutkimuksessa selvitettiin kanadalaisen Cognitive Skills -yleisohjelman vaikutuksia vankilaan palaamiseen (Honka ym., 2006)³. Lähtökohdaksi tutkimukselle oli se, että ohjelmien akkreditointimenettely edellytti tietoa ohjelmien uusintarikollisuusvaikutuksista. Tutkimusta voidaan pitää yhtäältä Rikosseuraamuslaitoksen pilottihankkeena, jonka tarkoituksena oli selvittää erilaisten ohjelmia koskevien lähdeaineistojen yhdistämistä keskusvankirekisterin tietoihin. Aineiston muodostaminen osoittautui hankalaksi. Ohjelmiin osallistuneille vangeille jouduttiin etsimään verrokkihenkilöt yksitellen keskusvankirekisteristä, eikä varsinaisia tilastollisia kaltaistamismenetelmiä käytetty. Kaltaistamismuuttujina

käytettiin sukupuolta, ikää, vankeuslajia, vankeuskertaisuutta, päärikosta, tuomion pituutta sekä vapautumisajankohtaa. Tutkimusaineiston koko oli 355 ohjelmaan osallistunutta ja 469 verrokkihenkilöä.

Ensimmäiset tutkimustulokset ohjelman vaikutuksista olivat lupaavia: ohjelman suorittaneiden vankien vankilaan palaamisen todennäköisyys oli pienempi kuin verrokkiryhmällä ja ohjelman keskeyttäneillä vangeilla. Kolmen seurantavuoden jälkeen ohjelman suorittaneista vankilaan oli palannut 48 prosenttia vangeista. Ohjelman keskeyttäneiden ja verrokkiryhmän uusimisosuus oli 61 prosenttia. Tulosten perusteella ohjelma ei soveltunut urautuneille vangeille; ohjelma vähensi vankilaan palaamista tilastollisesti merkittävästi ainoastaan ensimmäistä tai toista kertaa vapautuneilla vangeilla kontrolloitessa muiden taustamuuttujien vaikutukset. Tutkimuksessa tunnistettiin esimerkiksi valikoitumiseen liittyviä haasteita, jotka ovat vielä nykyäänkin ajankohtaisia (ks. luku *Vaiikutusten arvioinnin peruslähtökohdat*).

Cognitive Skills -ohjelman jatkotutkimuksessa (Tyni ym., 2014) ohjelman vaikutuksia arvioitiin aidolla uusintarikollisuudella. Tällä tarkoitetaan vankilaan palaamisen sijasta vapautumisen jälkeen tehtyjä rikoksia, jotka johtavat uuteen tuomioon. Suuri osa vangeista palaa vankilaan, mutta aina kyse ei ole rikoksista vapautumisen jälkeen. Jatkotutkimuksessa selvitettiin uusintarikollisuusvaikutusten ohella myös ohjelman välillisiä vai-

vaikutuksia vankien uusintarikollisuuteen; verrokkiryhmien uusiminen oli keskimäärin 16 prosenttia suurempaa kuin ohjelman aloittaneilla vangeilla. Tutkimustulosten yleistettävyyttä on vaikeuttanut mm. se, että tutkimusten kohderyhmät, tutkimusasetelmat, ohjelmien toteutustavat sekä vaikuttavuuden mittaaminen ovat eronneet huomattavasti toisistaan.

³ Cognitive Skills -ohjelman (Reasoning & Rehabilitation revised, R & R revised) vaikutuksista on tehty useita eri tutkimuksia (ks. esim. Wilkinson, 2005; Mitchell & Palmer, 2004; Berman, 2004; Cann ym., 2003; Ross ym., 1998). Esimerkiksi Tongin ja Farringtonin (2006) meta-analyysin perusteella ohjelmalla oli myönteisiä

kutuksia vankien motivaatioon ja ajattelutaitoihin puolistrukturoitujen alku- ja loppuhaastattelujen avulla. Verrokkiryhmä (N = 1743) muodostettiin matched pair -menetelmällä, jossa jokaiselle ohjelmaan osallistuneelle vangille (N = 672) pyrittiin poimimaan kolme täysin yhtenevää verrokkiparia iän, sukupuolen, pääríkoksen, laitossajan pituuden sekä vapautumisvuoden perusteella. Ohjelman todettiin lisäävän vankien motivaatiota ja ajattelutaitoja. Uusimisvaikutukset jäivät kuitenkin odotettua pienemmiksi: Cognitive Skills -ohjelman suorittaneista vangeista 54 prosenttia ja verrokkiryhmästä 59 prosenttia syyllistyi uusiin rikoksiin kolmen vuoden seuranta-aikana. Ohjelman keskeyttäneistä vangeista uusi peräti 70 prosenttia. Regressioanalyysien (logistinen regressioanalyysi ja Coxin suhteellisen vaaran malli) perusteella ohjelma ei kuitenkaan vähentänyt uusintarikollisuutta, kun muut uusimiseen vaikuttavat tekijät kontrolloitiin samanaikaisesti. Ohjelmalla todettiin sen sijaan olleen myönteisiä vaikutuksia motivaatioon ja ajattelutaitoihin.

Omaehtoisen muutoksen ohjelma

Omaehtoisen muutoksen ohjelma (OMA) on suunnattu toistuviin väkivaltarikoksiin syyllistyneille vangeille. Ohjelma on ryhmämuotoinen, 6–8 kuukautta kestävä pitkäkestoinen kuntoutusohjelma. Suomessa ohjelmasta on tehty kaksi vaikutustutkimusta⁴.

Ensimmäisessä tutkimuksessa (Konttila ym., 2011) aineistona olivat kaikki vuosina 2001–2006 ohjelman aloittaneet miesvangit (N = 110), jotka vapautuivat ennen vuotta 2008. Uusimisvaikutusten lisäksi tutkimuksessa

⁴ Kahden suomalaisen tutkimuksen lisäksi OMA-ohjelman uusintarikollisuusvaikutuksista on tehty ainoastaan yksi vertaisarvioitu tutkimus; Henningin ym. (1996) tutkimuksessa ohjelman läpikäyneet vangit syyllistyivät uusiin ri-

selvitettiin myös teoreettisesti relevanttien Rorschach Comprehensive System (RCS) -muuttujien (esim. tunteiden säätelykyky, voimakkaiden tunteiden hallinta ja minäkuva) sensitiivisyyttä ohjelman vaikutusten arvioinnissa pienellä otoksella (N = 6). Verrokkiryhmä muodostettiin lähes samoin kuin Cognitive Skills -tutkimuksessa. Verrokkipoolina oli toistuvista väkivaltarikoksista tuomitut pitkäaikaissangit. Kaltaistamismuuttujina käytettiin vankeuskertaisuutta, ikää, tuomion pituutta sekä vapautumisvuotta. OMA-ohjelman suorittaneista vangeista 68 prosenttia ja verrokeista 64 prosenttia syyllistyi uusiin rikoksiin kahden vuoden seuranta-aikana. Ohjelman keskeyttäneiden uusimisosuus oli odotetusti suurin (71 %). Toista, kolmatta ja neljättä kertaa vapautuneilla vangeilla ohjelmalla havaittiin viitteitä myönteisistä vaikutuksista, mutta erot ryhmien välillä eivät olleet tilastollisesti merkitseviä.

Asetelmaan liittyi ainakin kaksi ongelmaa. Verrokkiparien määrä jäi ensinnäkin odotettua pienemmäksi, koska suomalaisen vankipopulaation joukosta oli erittäin hankala löytää koeryhmälle kaikilta taustamuuttujiltaan täysin yhteensopivia vertaispareja. Joillekin ohjelmaan osallistuneille vangeille ei löytynyt yhtään verrokkiparia tai näitä löytyi yksi tai kaksi tavoitellun kolmen sijasta. Ohjelmaan osallistuneiden määrä toisin sanoen samalla pienensi voimakkaasti potentiaalisen verrokkipoolin kokoa. Toinen haaste liittyi osallistuneiden suhteelliseen pieneen määrään: aineiston koon takia uusimiserojen tuli ryhmien välillä olla suuria, jotta erot olisivat

koksiin tai vapautumisehtojen rikkomiseen harvemmin kuin verrokkihenkilöt. Tutkimustulosten yleistettävyyttä rajoittivat kuitenkin esimerkiksi aineiston pieneen kokoon ja kaltaistamismuuttujiin liittyvät puutteet.

tilastollisesti merkitseviä vaikutusten todentamiseksi (ks. luku *Vaikutusten arvioinnin peruslähtökohdat*).

OMA-ohjelmasta on tehty myös jatkotutkimus (Konttila ym. 2018). Jatkotutkimuksen tutkimusmenetelmällisenä tavoitteena oli laajentaa tutkimusaineiston kokoa sekä käyttää moderneja kaltaistamismenetelmiä verrokiparien valinnassa. Tutkimusaineiston muodostaminen pyrittiin tekemään suoravivaisesti, jäljitellen satunnaistetun kontrolloidun tutkimuksen toteuttamista. Ensimmäisessä vaiheessa Vankitietojärjestelmästä poimittiin kaikki ehdottomasta vankeudesta vapautuneet suomalaiset vangit, jotka olivat syyllistyneet vähintään yhteen väkivaltarikokseen. Tämän jälkeen aineistosta etsittiin kaikki OMA-ohjelmaan osallistuneet vangit sekä mahdollisimman suuri ryhmä muita vankeja, jotka olisivat rikostaustansa perusteella potentiaalisesti voineet osallistua OMA-ohjelmaan (verrokipooli). Seuraavaksi poimittiin verrokkihenkilöt. Kaltaistamismuuttujina käytettiin tietoja henkilöiden iästä, sukupuolesta, aikaisemmasta väkivaltarikollisuudesta (henkirikokset, henkirikosten yritykset, pahoinpitelyt), vankeuskertaisuudesta, nykyisen vankeuskauden pituudesta sekä vapautumisvuodesta.

Kaltaistaminen tehtiin käyttämällä R-ohjelmiston Matching-kirjaston Genetic Matching-algoritmia (Diamond & Sekhon, 2013) ilman verrokkihenkilöiden takaisinpanoa. Kaltaistamisanalyysien perusteella kävi ilmi, että kahden verrokin poimiminen kullekin ohjelman aloittaneelle vangille tuotti paremman taustamuuttujien tasapainon koe- ja verrokkiryhmien välillä kuin alun perin suunniteltu kolmen verrokkihenkilön poiminta. Lopulliseen aineistoon valittiin 109 ohjelman aloittanutta ja 218 verrokkia. Tiedot vankien uusintarikollisuudesta liitettiin aineistoon vasta tämän jälkeen. Tutkimuksessa analysoitiin

sekä uusintarikollisuutta väkivaltarikoksiin että mihin tahansa muihin rikoksiin. OMA-ohjelman tulokset jäivät valitettavan heikoiksi. Yhden vuoden seuranta-aikana 48 prosenttia ohjelman aloittaneista ja 44 prosenttia verrokkiryhmästä uusi minkä tahansa rikoksen. Ryhmien välinen ero (HR = 1,10; CI 95 % = 0,81–1,49) ei kuitenkaan ollut tilastollisesti merkitsevä; Ohjelman suorittaneista uusiin rikoksiin syyllistyi 45 prosenttia (HR = 1,04; CI 95 % = 0,75–1,43). Väkivaltarikoksiin syyllistyi yhden vuoden seuranta-aikana ohjelmaan osallistuneista 27 prosenttia (HR = 1,24; CI 95 % = 0,86–1,80) ja verrokeista 17 prosenttia. OMA-tutkimus on ollut toistaiseksi viimeisin ohjelmien uusintarikollisuusvaikutuksia selvittävä tutkimus.

STOP

Seksuaalirikollisille suunnattu STOP-ohjelma oli käytössä Suomessa vuosina 2000–2018. Ohjelmasta on tehty yksi vaikutustutkimus (Laaksonen & Tyni, 2015). Tutkimuksessa selvitettiin aluksi millaiset staattiset taustatekijät ennustavat seksuaalirikoksen uusimista suomalaisilla seksuaalirikosvangeilla (N = 693). Seuraavaksi ohjelman suorittaneille vangeille (N = 143) poimittiin seksuaalirikollisista tuomittujen vankien verrokipoolista verrokkihenkilöt (N = 143). Kaltaistamismenetelmänä käytettiin logistiseen regressioanalyysiin perustuvaa propensiteetipistemäärämenetelmää ("propensity score matching", Austin, 2011), jossa aluksi estimoitiin todennäköisyys osallistua ohjelmaan eri taustatekijöiden valossa. Tätä pistemäärää hyödyntämällä voitiin etsiä kullekin koeryhmän henkilölle verrokki, jolla oli taustatekijöiden valossa yhtä suuri todennäköisyys osallistua ohjelmaan. STOP-ohjelmaan osallistuneista vangeista 6 prosenttia ja verrokki-

ryhmästä 11 prosenttia syyllistyi uusiin seksuaalirikoksiin tutkimuksen seuranta-aikana (keskimäärin 7,5 vuotta). Ryhmien välinen ero ei kuitenkaan ollut tilastollisesti merkitsevää⁵. Ohjelma ei ole enää käytössä Suomessa, vaan se on korvattu toisella ohjelmalla (STEP). STEP-ohjelman vaikuttavuutta ei ole arvioitu.

Rikosseuraamuslaitoksen toiminnan uusimisvaikutuksia on arvioitu myös muiden toimenpiteiden näkökulmista, jotka eivät ole liittyneet uusintarikollisuuteen vaikuttaviin kuntoutusohjelmiin. Tutkimusmenetelmällisesti kyseiset tutkimukset ovat vertailukelpoisia edellä mainittuihin monista eri lähtökohdista. Vankilakoulutuksen vaikutuksia arvioineessa tutkimuksessa (Virtanen ym., 2020) ei havaittu uusintarikollisuuseroja tutkintoja suorittaneiden vankien ja heidän verrokkiensa välillä, mutta koulutuksen suorittaneiden työmarkkinatilanne oli vapautumisen jälkeen jonkin verran verrokkeja parempi erityisesti vähintään toisen asteen tutkinnon suorittaneiden joukossa.⁶

VAIKUTUSTEN ARVIOINNIN PERUSLÄHTÖKOHDAT

Jotta voimme sanoa mitään mielekästä ohjelmien kausaalivaikutuksesta uusintarikollisuuteen, tarvitsemme ”kontrafaktuaalisen”

⁵ Mewsin ym. (2017) tutkimuksen perusteella ohjelma (Core SOTP) ei vähentänyt seksuaalirikoksista tuomittujen vankien uusintarikollisuutta. Uusintarikollisuusvaikutukset olivat lähinnä kontraproduktiivisia eli ohjelman läpikäyneet vangit uusivat keskimäärin useammin kuin verrokkit. Tutkimuksessa käytettiin poikkeuksellisen suurta koeryhmää (N = 2 562) ja verrokkiryhmää (N = 13 219), pitkää seuranta-aikaa (keskimäärin 8,2 vuotta) sekä useita kaltaistamismuuttujia (87). Myös Friendshipin ym. (2003) tutkimustulosten perusteella ohjelman vaikutukset olivat vaatimattomia.

arvion siitä, kuinka yleistä koeryhmän uusintarikollisuus olisi ollut ilman ohjelmaan osallistumista. Yksinkertaisimmillaan kyse on toisin sanoen kahden *vaihtoehtoisen lopputuleman* (uusintarikollisuus ohjelman läpikäyneenä vs. ilman ohjelmaa) eroa yksilötasolla. Koska näitä vaihtoehtoisia todellisuuksia ei voida yksilötasolla havaita ja vertailla, päättelyn on tapahduttava ryhmätasolla. Kausaalivaikutuksista kiinnostuneet tutkijat pyrkivät laskemaan uskottavimman mahdollisen kontrafaktuaalisen arvion kyseiseen ohjelmaan osallistumattoman verrokkiryhmän avulla, ja vertaamaan tätä koeryhmän toteutuneeseen uusintarikollisuuteen. Mikäli uskottavan verrokkiryhmän muodostaminen onnistuu (=vaihtoehtoisissa lopputulemissa ei ole eroa koe- ja verrokkiryhmien välillä), voimme estimoida ohjelman keskimääräisen vaikutuksen uusintarikollisuuteen koe- ja verrokkiryhmää vertaamalla. Suomessa vaikutusarviointia on tämän jälkeen mahdollista laajentaa rekisteritietojen avulla myös muihin lopputulemiin, kuten työssäkäyntiin (Virtanen ym., 2020), tuloihin tai vaikkapa sairastavuuteen. Subjektiviisempia kokemuksia esimerkiksi ajattelutapojen muutoksista rekisteriaineistot eivät mittaa, joten näiden vaikutusten tarkasteluun tarvitaan muita aineistonkeruutapoja.

⁶ Lisäksi on esim. tutkittu motivoivaan haastateluun perustuvan Move! -ohjelman vaikuttavuutta (Väyrynen ym., 2014) ja kuntouttavan vankityön uusimisvaikutuksia (Virta, 2013), vertailtu törkeästä rattijuopumuksesta ehdottomaan vankeuteen ja yhdyskuntapalvelurangaistukseen tuomittujen uusintarikollisuutta (Sirén, 2012) sekä selvitetty päihdehoidossa olleiden laitos- ja avohoitopotilaiden vankeuteen johtanutta rikollisuutta (Pitkänen ym., 2016). Viimeisinä vuosina on tehty myös joitakin vankilateatteriin liittyviä tutkimuksia (ks. Menard, 2018).

Uskottavan verrokkiryhmän löytäminen on kuitenkin valitettavan haastava tehtävä. Perusongelmana kaikessa havainnoivaan tutkimusasetelmaan (vrt. satunnaiset kontrolloidut asetelmat) perustuvassa vaikutusarvioinnissa on se, että ohjelmiin osallistuminen ei ole lähtökohtaisesti satunnaista. Tutkijoiden on vaikea jälkikäteisesti selvittää, millä perusteella tietyt henkilöt osallistuivat ohjelmiin. Ongelma pahenee, kun ohjelmiin osallistuminen on vapaaehtoista. Vaikka löytäisimmekin potentiaalisten verrokkien joukosta vankeja, jotka vastaavat mitattavissa olevien taustamuuttujien valossa koeryhmää, jäljelle jää kysymys siitä, miksi näistä ”tilastollisesti samankaltaisista” henkilöistä yksi osallistui ohjelmaan, mutta toinen ei? On varsin todennäköistä, että esimerkiksi vankeuskertaisuudeltaan toisiaan vastaavien vankien joukossa on vielä huomattavaa vaihtelua uusintarikollisuuden riskissä ja siihen vaikuttavissa taustatekijöissä, mutta emme pysty mittaamaan tätä riskiä käytettävissä olevan aineiston avulla kovinkaan luotettavasti. Tästä havaitsemattomien muuttujien ongelmasta seuraa, että sekä regressiomalleja että kaltaistutusmenetelmiä hyödyntävät, havaittujen muuttujien vakiointiin tai tasapainottamiseen perustuvat tutkimukset ovat ongelmallisia vaikutusten arvioinnissa. Kaikki Suomessa toistaiseksi toteutetut ohjelmia koskevat arvioinnit ovat perustuneet tällaisiin tutkimusasetelmiin.

Aika usein erilaisten yksilötasoisten interventioiden arvioinnissa oletetaan, että halu irtautua rikosuralta, desistanssimotivaatio, ennustaisi sekä ohjelmiin osallistumista että niiden onnistunutta suorittamista (Bushway & Apel, 2012). Mikäli näin on, ja desistanssi-

motivaatiolla on itsenäinen vaikutus uusintarikollisuuteen, ovat ohjelmiin osallistuvat vangit ”positiivisesti valikoituneita” jo lähtötilanteessa ennen ohjelman alkua. Ongelma korostuu, jos mukaan otetaan ainoastaan ohjelman onnistuneesti suorittaneet vangit, ja keskeyttäneet jätetään analyysin ulkopuolelle⁷. Tällaisessa tilanteessa rikostaustan vakiointi tuskin yksin riittää tavoittamaan motivaatioeroja, ja tutkimusasetelma vinoutuu ohjelmalle edulliseksi, vaikkei ohjelmalla olisikaan todellista vaikutusta uusintarikollisuuteen. On kuitenkin vaikeampi sanoa, missä määrin tämä oletus positiivisesta valikoitumisesta ohjelmiin pätee suomalaisessa vankipopulaatiossa. Edellä esitetyistä tilastoista selviää, että ohjelmia suorittaneet vangit ovat tyypillisesti olleet vankilassa jo useita kertoja, ja heidän vankeuskautensa ovat huomattavasti ohjelmia suorittamattomia vankeja pidempää.

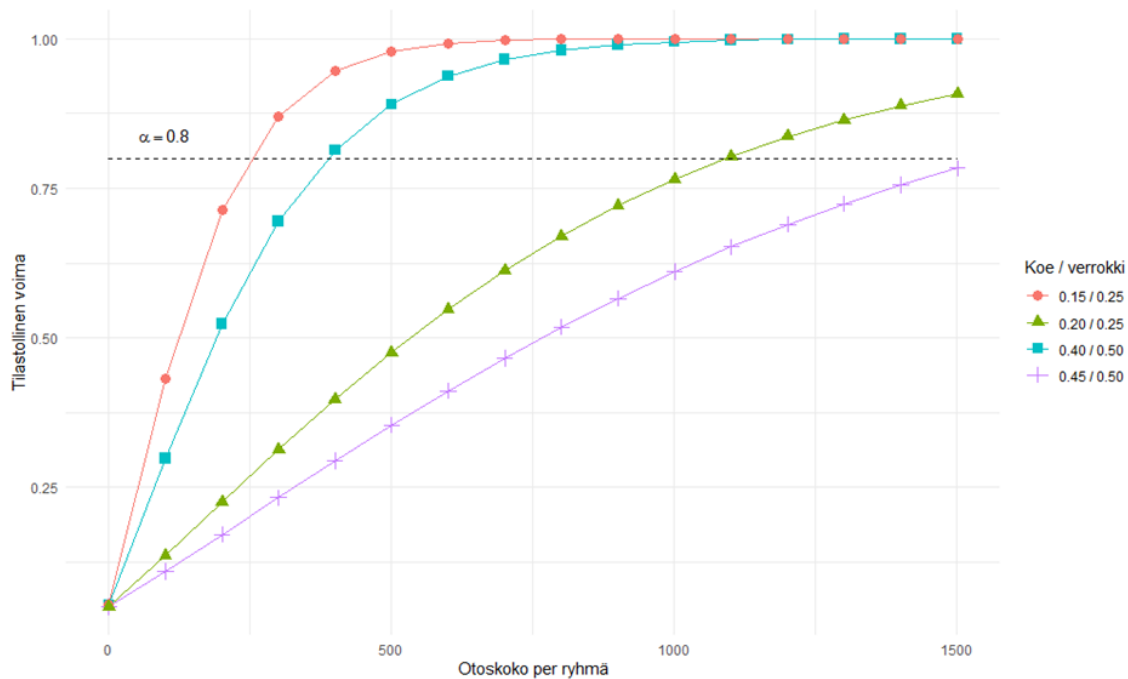
Valikoitumisongelmasta ja tavoista sen ratkaisemiseen satunnaistettujen ja luonnollisten koeasetelmien avulla on paljon tuoretta kirjallisuutta (Angrist & Pischke, 2014; Morgan & Winship, 2014) ja myös hyviä kotimaisia yleisesityksiä (Nokso-Koivisto ym., 2019). Valikoitumisongelman sijasta keskitymme seuraavassa eräisiin muihin vaikutustutkimusta hankaloittaviin seikkoihin, jotka vaikuttavuuden mittaamisessa olisi jatkossa hyvä huomioida.

Tilastollinen voima

Olellainen osa satunnaistetun kontrolloidun kokeen suunnittelua on tutkimusasetelman tilastollisen voiman arviointi *odotettavissa*

⁷ Suomalaisissakin vaikutusarvioinneissa (esim. Konttila ym., 2018) on havaittu keskeyttäneiden uusivan ohjelman suorittaneita enemmän. Tämä

havainto ei sellaisenaan kuitenkaan todista ohjelman vaikutuksia, sillä keskeyttäneiden ja suorittaneiden välillä voi olla systemaattisia eroja.



Kuvio 3. Tilastollinen voima otoskoon sekä koe- ja verrokkiryhmien uusimisasteen mukaan.

olevien vaikutusestimaattien valossa. Jos tutkimusta suunnitteleva tutkija olettaa aikaisempaan tutkimukseen, substanssialueen vahvaan tuntemukseen tai ”yleiseen elämäkokemukseen” perustuen, että ohjelman keskimääräinen vaikutus uusintarikollisuusasteeseen (=koe- ja verrokkiryhmien ero rikosten uusineiden osuudessa) on luultavasti pieni, hänen tulisi pyrkiä kasvattamaan otoksen kokoa. Liian pienen otoksen käyttö on pahimmillaan sekä tutkimukseen osallistuvien henkilöiden että tutkijoiden ajan haaskausta ja jopa tutkimuseettisesti kestämättömää, sillä näillä asetelmilla saavutettavat estimaatit ovat jo lähtökohtaisesti liian epätarkkoja.

Otoskoon riittävyttä voidaan arvioida tilastollisen voiman ja voimakkuuslaskelmien avulla jo ennen aineiston keruuta. Koska käytämme esimerkissämme dikotomista vastemuuttujaa, kuvaamme tässä voimakkuuslaskelman teon kahden suhteellisen osuuden

testin näkökulmasta. Voimakkuuslaskelman tekemistä varten tarvitsemme tiedon oletetusta vaikutuksen suuruudesta (paras arvaus uusineiden osuudesta koe- ja verrokkiryhmissä, ts. ohjelman todellinen vaikutus), valitusta merkitsevyystasosta β ja tavoitellusta tilastollisesta voimasta α . Kuten tilastollisessa testaamisessa yleensä, β :n tyypillisin arvo on 0.05. Tilastollisen voiman raja-arvona α käytetään yleisesti arvoa 0.80, mutta se voi olla korkeampikin. Kun nämä arvot on päätetty, voimme laskea tarvittavan otoskoon n tilanteessa, jossa haluamme havaita tilastollisesti merkitsevän eron ($p < 0.05$) kahden ryhmän välillä 80 % todennäköisyydellä. Kääntäen tilastollinen voima kertoo, kuinka todennäköisesti havaitsemme otoksessamme tilastollisesti merkitsevän eron koe- ja verrokkiryhmien välillä, kun todellinen vaikutus on oletetun suuruinen ja otoskoko on n (Cohen, 1988).⁸ Mitä pienemmäksi oletamme ryhmien välisen eron, sitä suuremman aineiston

⁸ Suomenkielinen johdatus tilastollisen voiman laskentaan: <http://www.reito.fi/tilastokunto/tilastollinen-voima>

tarvitsemme sen luotettavaan havaitsemi-
seen⁹.

Havainnollistamme voimakkuuslaskelmaa seuraavassa yksinkertaisen esimerkin avulla. Kuvion 3 neljä viivaa kuvaavat tilastollisen voiman muutosta aineiston koon kasvattamisen myötä neljässä eri skenaariossa, jossa koe- ja verrokkiryhmien todellinen uusimisen ero (=ohjelman vaikutus) on joko viisi tai kymmenen prosenttiyksikköä, ja verrokkiryhmän uusimisaste on joko 50 % tai 25 %. Laskelmassa oletetaan, että koe- ja verrokkiryhmät ovat yhtä suuret. Kuvion x-akseli kuvaa yhden ryhmän otoskokoa, y-akseli tilastollista voima. Kuvion vaakaviiva kuvaa valittua tilastollisen voiman raja-arvoa $\alpha=0.80$. Laskelma tehtiin R-ohjelmiston kirjastolla ”pwr” (Champely ym., 2020).

Laskelma osoittaa, että tilanne, jossa verrokkiryhmän todellinen uusimisaste on 25 %, ja koeryhmän 15 % (ohjelman suorittaminen vähentää uusimista kymmenellä prosenttiyksiköllä), on tutkimuksen kannalta otollisin. Tilastollinen voima $\alpha=0.80$ saavutetaan, kun koe- ja verrokkiryhmiin kuuluu noin 250 henkilöä, eli aineiston yhteenlaskettu koko on noin 500 henkilöä. Tilanne, jossa uusimisen ero on prosenttiyksiköissä sama, mutta uusimisaste absoluuttisesti korkeampi, vaatii jo selvästi suuremman aineiston, noin 750 henkilöä. Jos sen sijaan oletamme, että ohjelman vaikutus uusimiseen olisi viiden prosenttiyksikön luokkaa, ja verrokkiryhmästä uusisi joka toinen, tilanne on tutkijan kannalta jo huomattavasti hankalampi. Saavuttaaksemme tavoitellun tilastollisen voiman, aineiston koon pitäisi olla yli 3000 henkilöä, mikä vastaa jo yhtenä vuotena vapautuvien vankeusvankien kokonaismäärää.

⁹ On kokonaan toinen kysymys, kannattaako tilastolliseen päättelyyn suhtautua ”binäärisesti”

Tutkimusaineistojen laatu ja luotettavuus

Tutkimusaineistojen osalta vähimmäisvaatimuksena yksilötason analyysiin perustuvalla vaikutustutkimukselle on tieto siitä, ketkä vangit ohjelmiin osallistuivat. Lisäksi kaivataan ainakin erittely ohjelman keskeyttäneisiin ja suorittaneisiin. Ohjelmatyön alkuvaiheessa 1990-luvun lopussa ja 2000-luvun alussa vankiloiden ohjaajat keräsivät tietoa kuntoutusohjelmiin osallistuneista vangeista erillisiksi tiedostoiksi, jotka lähetettiin Rikosseuraamuslaitoksen (Rikosseuraamusviraston) keskushallintoyksikköön (Honka ym., 2006; Tyni ym., 2014). Tietotarpeet liittyivät lähinnä kuntoutusohjelmien osallistumismäärien seurantaan ja tilastointiin eivätkä niinkään suoranaisesti ohjelmien vaikuttavuuden mittaamiseen. Laajamittaisten ryhmämuotoisten akkreditoitujen kuntoutusohjelmien osalta tilanne oli parempi, koska tiedot kerättiin ennalta suunniteltuihin lomakkeisiin yhteisillä kirjaamiskäytännöillä.

Vuosina 2005–2006 Rikosseuraamuslaitos otti asteittain käyttöön Vankitietojärjestelmän, jonka jälkeen vankiloiden ohjaajia ohjeistettiin kirjaamaan kaikki ohjelmiin osallistuneet vangit järjestelmään. Vankitietojärjestelmän ohjelmaosuuteen (ohjelmanäyttö) kirjataan tällä hetkellä kullekin vangille ohjelman aloittamisaika, suorittamisaika sekä mahdollinen keskeyttämispäivä ja keskeytyksen syy. Vankitietojärjestelmän käyttöönoton jälkeen oleellista vaikutustutkimuksen näkökulmasta oli myös se, että osallistumistietoihin oli mahdollista liittää tietoja muun muassa vankien rikoshistoriasta ja ohjelman ja keskittyä sinänsä sopimuksenvaraisiin p-arvojen rajoihin ohjelmien vaikuttavuuden mittarina (Gelman ym., 2017).

jälkeisestä uusintarikollisuudesta joko suoraan Rikosseuraamuslaitoksen rekistereistä tai ulkopuolisista rekisterilähteistä asianmukaisilla tutkimusluvilla. Tällaisella taustatiedolla voidaan paremmin ottaa huomioon koe- ja verrokkiryhmien välisiä eroja, ja toisaalta laajentaa seuranta pelkkien uusintarikollisuutta koskevia muuttujia monipuolisemmaksi (Hyatt & Andersen, 2019).

Uusien ohjelmien aloittaminen tai vanhojen ohjelmien poistaminen vankiloiden ohjelma- kirjosta saattaa vaikuttaa nopeasti ohjelmien suoritemääriin. Lisäksi Rikosseuraamuslaitoksen kirjaamisohjeistuksilla voi olla suora vaikutus siihen, millaisia ohjelmia ohjaajat ovat kirjanneet Vankitietojärjestelmän ohjelmaosuuteen. Suurin vaikutustutkimuksien haaste on ollut se, että kaikkia ohjelman suorittaneita vankeja ei ohjeistuksesta huolimatta ole viety Vankitietojärjestelmään. Akkreditoitujen ohjelmien osalta tiedot ovat pääsääntöisesti kirjattu mallikkaasti, mutta etenkin joissakin lyhyissä ohjelmissa puuttuvien tietojen osuus on ollut suuri, mikä vaikeuttaa koeryhmän jäsenten identifiointia. Puuttuvien tietojen ongelma koskee tietysti myös muita Vankitietojärjestelmän kirjauksia, joita analyysissä halutaan käyttää (Tyni, 2015). Tästä syystä luotettava ja johdonmukainen kirjaaminen on keskeinen edellytys vaikutustutkimukselle.

Verrokkiryhmän vankeusajan sisältö ja ohjelmien räätälöinti

Kaltaistetussa tutkimusasetelmassa, jossa tiettyyn ohjelmaan osallistuvien henkilöiden verrokkit poimitaan muiden vankien joukosta, ongelmaksi saattaa lisäksi muodostua se, ettei ole selvää, millaiseen vankeuden sisältöön vertailu milloinkin tapahtuu. Tämä on-

gelma liittyy pitkälti edellä mainittuun kirjausten kattavuutta koskevaan kysymykseen. On todennäköistä, että Suomessa tehdyissä edellä kuvatuissa arvioinneissa verrokkiryhmän jäsenet ovat myös osallistuneet erilaisiin toimintamuotoihin vankeuden aikana, ja mikäli näillä on myönteisiä vaikutuksia uusintarikollisuuteen, saattaa sinänsä hyödyllisen ohjelman vaikutus jäädä piiloon. Verrokkiryhmän valintaa koskeva ongelma ei siten koske pelkästään valikoitumista vankeutta edeltävien taustatekijöiden suhteen, vaan myös vankeuden erilaisen sisällön suhteen. Tätä ongelmaa voidaan osin torjua rajaamalla potentiaalisten verrokkien ryhmää mitattavissa olevien, vankeusaikaa koskevien tietojen avulla, mutta aivan helppoa se ei ole.

Ensimmäinen pohdittava seikka on koeryhmän muu vankeusaikainen toiminta. Esimerkiksi tätä artikkelia varten koostetussa vuosina 2010–2019 vapautuneiden vankien tutkimusaineistossa uusintarikollisuuteen vaikuttavan ohjelman suoritti 14 prosenttia kaikista vuonna 2019 vapautuneista vankeusvangeista (N = 3138). Suuri osa tästä 14 prosentista suoritti uusintarikollisuuteen vaikuttavien ohjelmien lisäksi myös päihdekuntoutusohjelman (12,3 %), sosiaalisen kuntoutusohjelman (14,1 %) tai molemmat näistä (10,1 %) vankeusaikana. Kuntoutusohjelmien lisäksi vangit saattavat osallistua vankityöhön, koulutukseen tai johonkin muuhun vankilan hyväksymään toimintamuotoon vankeutensa aikana. Toimintamuotoihin osallistumiseen ovat yhteydessä muun muassa rangaistuksen pituus, toimintovalikoima suljetuissa vankiloissa ja avovankiloissa sekä rangaistusajan suunnitelmaan kirjatut tavoitteet.

Verrokkien vankeusajan sisältöä voidaan tarkastella esimerkkitapauksen kautta. Viimeiseksi yhdistimme Vankitietojärjestelmästä poimitun perusaineiston (N = 27 726)

väkivaltarikollisille suunnatun OMA-tutkimuksen (ks. Konttila ym., 2018) tutkimusaineistoon. OMA-tutkimuksessa aineistona oli 109 ohjelmassa mukana ollutta (suorittaneet ja keskeyttäneet) vankia ja 218 kaltaistettua verrokkiä. Kaltaistamismuuttujina OMA-tutkimuksessa käytettiin ikää, sukupuolta, tietoa aikaisemmasta väkivaltarikollisuudesta, tuomion pituutta ja vapautumisvuotta. Konttilan ym. (2018) tutkimuksessa kontrolloitiin lisäksi joltain osin kuntoutusohjelmiin osallistumista. Verrokkihenkilöt eivät esimerkiksi voineet osallistua muihin uusintarikollisuuteen vaikuttaviin pidempikestoisiin kuntoutusohjelmiin.

Artikkelimme tutkimusaineistoon kiinnittyi 96 OMA-ohjelman suorittanutta vankia ja heidän 178 kaltaistettua verrokkiään. Vangit olivat vapautuneet vuosien 2010–2015 välisenä aikana. Aineiston perustella OMA-ohjelman suorittaneet ja kaltaistetut verrokki osallistuivat vankeutensa aikana usein myös muihin kuntoutusohjelmiin. OMA-ohjelman läpikäyneistä vangeista joka kolmas (33 %) oli suorittanut vankilassa jonkin päihdeohjelman ja lähes joka kymmenellä (8 %) oli suoritusmerkintä sosiaalisen kuntoutuksen ohjelmasta; koeryhmän vangeista 7 prosenttia oli lisäksi läpikäynyt sekä päihdeohjelman että sosiaalisen kuntoutusohjelman. Kaltaistetuista verrokeista peräti joka toinen (52 %) oli suorittanut uusintarikollisuuteen vaikuttavan ohjelman. Lisäksi joka toisella vangilla oli suoritusmerkintä päihdekuntoutusohjelmasta (51 %) ja joka seitsemännellä (15 %) muusta sosiaalisesta kuntoutuksesta. Suurin osa verrokkiryhmän vankien uusintarikollisuuteen vaikuttavista ohjelmista oli kirjattu joko motiivintehohjelmille (lähinnä Viisi Keskustelua Muutoksesta) tai lyhyemmille vaikuttavuusohjelmille (mm. Suuttumuksen hallinta). On todennäköistä, että tämä OMA-ohjelman aineiston avulla havainnollistamamme tilanne

haittaa muidenkin ohjelmien vaikutusten arvioimista: suuri osa niin koe- kuin verrokki-ryhmän jäsenistä osallistuu vankeuden aikana muihinkin kuin tutkittavaan interventioon.

Rikosseuraamuslaitoksessa on käyty keskustelua esimerkiksi siitä, mihin vaiheeseen kuntoutusohjelmat tulisi ajoittaa rangaistusajankana. Pitkäaikaisvankien tapauksessa pitkäkestoiset ja intensiivisimmät ohjelmat suoritetaan yleensä suljetussa vankilassa, jota edeltää mahdollinen päihdekuntoutusjakso. Myöhemmin vanki pyritään sijoittamaan avovankilaan ja mahdollisen koevapauden kautta vapauteen. Lyhytaikaisvangit voidaan puolestaan sijoittaa suoraan avovankilaan, joiden tarjonta keskittyy lyhyisiin ja keskipitkiin kuntoutusohjelmiin. Keskeiseksi kysymykseksi nousee se, tuleeko opittuja malleja harjoitella ensin vankilassa vai ajoittaa ohjelman suorittaminen lähelle vapautumista. Toinen keskeinen kysymys on se, millaisille vangeille tietty ohjelmatyyppi sopisi parhaiten, ja onko kuntoutuksen teho parempi jossain alaryhmässä. Kysymykset ajoituksesta ja ohjelmien kohdentamisesta ovat tärkeitä, mutta tutkimuksen kannalta ne asettavat lisähaasteita. Jos ohjelman suorittaneiden nykyisillä määrillä ei onnistuta luotettavasti estimoimaan edes keskimääräisiä vaikutuksia, tämän joukon osittaminen useampaan alaryhmään (varioiden esimerkiksi ohjelman aloitusajankohtaa) vähentää tilastollista voimaa entisestään.

JOHTOPÄÄTÖKSET

Tarve vahvemmille tutkimusasetelmille

Tekemämme tutkimuskatsaus osoittaa, että olemme ohjelmatoiminnan vaikuttavuuden arvioinnin osalta edelleen (Tyni, 2015) siinä tilanteessa, ettei selvää näyttöä yhdenkään

ohjelman uusintarikollisuusvaikutuksista ole löydetty. On selvää, että vankeudenaikaisten kuntoutusohjelmien vaikuttavuuden tilastollisessa arvioinnissa on vielä paljon parannettavaa menetelmällisesti. Kaikki julkaistut tutkimukset perustuvat havainnoiviin tutkimusasetelmiin, ja pääosin jonkinlaisen kaltaistetun verrokkiryhmän käyttöön. On epätodennäköistä, että näillä tutkimusasetelmilla voitaisiin ottaa huomioon kaikkia koe- ja verrokkiryhmien välisiä eroja. Julkaistujen tutkimusten ”harhan” suunnasta ei myöskään voi olla täysin varma. Raportoidun tutkimusevidenssin valossa vaikuttaa kuitenkin mielestämme siltä, että ohjelmien vaikutukset ovat todennäköisesti verrattain pieniä. Joka tapauksessa arviointitutkimusten perusteella tiedämme, että merkittävä osa ohjelmia suorittaneistakin syyllysty uusiin rikoksiin ja palaa takaisin vankilaan.

Tutkimusaineistomme perusteella havaitsimme myös eräitä keskeisiä puutteita ohjelmien toteutuksessa sekä niitä koskevien järjestelmäkirjausten saatavuudessa ja luotettavuudessa. Esimerkiksi ajankäyttötietojen tilastokategoriat muuttuivat useamman kerran seuranta-aikana. Vaikuttavuustutkimuksen viitekehyksessä arvioinnin luotettavuutta lisää se, että ohjelmat ovat selvästi määriteltyjä sisällöltään, sekä toteutettu aiotulla tavalla ja riittävässä laajuudessa. Muuten emme voi tietää, millaista toimintaa olemme oikeastaan arvioimassa. Tutkimuksen empiirisen analyysin toteutuksen osalta vähimmäisvaatimuksena on puolestaan se, että vankien tausta- ja osallistumistiedot ovat kirjattu kattavasti tietojärjestelmään.

Jälkikäteen arvioiden voidaan todeta, että Suomessa tehtyjen arviointitutkimusten aineistojen koot ovat vielä pääsääntöisesti jääneet liian pieniksi. Uusien vaikutustutkimusten suunnittelun näkökulmasta tämä tarkoittaa sitä, että nykyisillä yksittäisten ohjelmien

osallistujamäärillä on vaikea tehdä tutkimusta, jossa tilastollinen voima olisi riittävä nyt realistiselta tuntuvan uusintarikollisuusvaikutuksen havaitsemiseen. Vaikka ohjelmien optimaalisesta ajoituksesta ja oikeasta kohdentamisesta olisi ehdottomasti syytä tehdä tutkimusta, tällä hetkellä olemme tilanteessa, jossa luotettava tieto keskimääräisistäkin vaikutuksista olisi merkittävä lisä tietopohjaan. Sama haaste koskee kansainvälisen tutkimusevidenssin hyödyntämistä kotimaisten ohjelmien suunnittelussa. Siitä huolimatta, että ulkomailta on saatavilla korkeatasoisempiin tutkimusasetelmiin perustuvaa näyttöä erilaisten interventoiden vaikutuksista, vaivaavat tässä artikkelissa kuvatut haasteet myös merkittävää osaa kansainvälisestä tutkimuksesta. Olisikin tärkeää, että ulkomailta menestyksekkäiksi väitettyjen toimintamuotojen arvioinnit käytäisiin tarkasti läpi menetelmällisen laadun näkökulmasta.

Ohjelmien pienetkin uusintarikollisuutta vähentävät vaikutukset puoltavat jo sellaisenaan niiden käyttämistä vankeinhoidossa. Uusintarikollisuutta onnistuneesti vähentävät ohjelmat ovat lisäksi potentiaalisesti erittäin kustannustehokkaita pelkästään vankeuden korkean hinnan vuoksi: esimerkiksi vuonna 2019 yhden vankilavuoden arvioitu hinta suljetussa vankilassa oli noin 78 000 euroa ja avolaitoksessa noin 54 000 euroa (Rikosseuraamuslaitos B). Todellisuudessa vankien rikollisuudesta aiheutuneet kustannukset ovat vielä korkeampia, koska vankilavuoden hinta ei sisällä rikoksen tutkinnasta, oikeudenkäynnistä tai rikoksen uhrille (ja läheisille) aiheutuneita euromääräisiä tai mittaamattomia kustannuksia. Tästä syystä pienetkin uusintarikollisuusvaikutukset olisi tärkeää saada luotettavasti esille. Lisäksi ohjelmatoiminta tarjoaa vangeille mielekäästä toimintaa vankilassa: ohjelmat voivat lisätä vankien muita valmiuksia rikoksettomaan elämään,

vaikka todellisia uusimisvaikutuksia ei havaittaisikaan.

Ohjelmatoiminnan tulevaisuus?

Artikkelimme tulokset antavat jossain määrin ristiriitaisen kuvan vankiloiden ohjelmatoiminnan tilasta. Mikäli tarkastelemme vapautuneita vankeja, näyttää siltä, että ohjelmia suorittaneiden vankien osuus vapautuneista vangeista on kasvanut 2010-luvulla tasaisesti, ja ohjelmatoiminta olisi siis tällä mittarilla laajentunut. Toisaalta kuitenkin havaitsimme, että akkreditoitujen ohjelmien suorittemäärät ovat laskeneet samalla tarkastelujaksolla selvästi. Tarkempi ohjelmatyyppien tarkastelu osoittaa, että ohjelmatoiminnan laajenemista selittää ennen kaikkea pääasiassa lyhytkestoisten sosiaalisten kuntoutusohjelmien suorittemäärien nousu sekä Viisi Keskustelua Muutoksesta -ohjelman suuret osallistujamäärät. Rikosseuraamuslaitoksen ohjelmatoiminnan pääpaino onkin näiden aineistojen perusteella nykyisin entistä vahvemmin lyhyillä kuntoutusohjelmilla.

Aineistomme perusteella pitkäkestoisten kuntoutusohjelmien tilanne näyttää huolestuttavalta. Säästöpainneessakin ohjelmatoiminnalle olisi varattava riittävät resurssit ohjelmien toteuttamiseksi. Resursseja tulisi kohdentaa erityisesti ohjaajien perus- ja jatkokoulutukseen ja metoditukeen sekä säilyttää ohjelmatoiminnan asema vankiloiden kuntoutustarjonnassa.

Vankiloiden ohjelmatoiminta näyttäytyi havaintojemme mukaan hyvin pirstaleisena. Laaja ohjelmakirjo tai suuret ohjelmien suorittemäärät eivät välttämättä tarkoita korkeaa laatua. Vuonna 2019 Vankitietojärjestelmään kirjattiin suorituksia yhteensä 37 eri ohjelmalle, ja tässäkin lukumäärässä on mukana useampia kaatoluokkia, mikä tarkoittaa, että

todellinen ohjelmien kirjo on tätäkin suurempi. Voidaan kysyä, olisiko sittenkin vaikuttavampaa keskittää resurssit suppeampaan, mutta näyttöperustaltaan vahvempaan ohjelmatoimintaan? Ohjelmatoiminnan linjauksissa määriteltyihin toimintatapoihin ja ohjelmiin tulisi sitoutua kaikissa Rikosseuraamuslaitoksen yksiköissä aikaisempaa paremmin sekä laajentaa hyväksyttämismenettely tiiviimmin osaksi ohjelmien valinta- ja seurantaprosessia. Parhaillaan käynnissä oleva Rikosseuraamuslaitoksen organisaatio-uudistus voisi tarjota mahdollisuuden ohjelmatoiminnan kehittämiseen.

Lopuksi

Ohjelmien pienistä vaikutuksista ja tilastollisen voiman puutteesta seuraa ongelmia myös rikosentorjuntaohjelmien rahoituksesta päättävälle taholle. Niukkoja resursseja olisi syytä käyttää toimenpiteisiin, joilla on todistetusti rikollisuutta vähentävä tai muu myönteinen vaikutus ohjelman kohderyhmään. Laajemminkin ajateltuna suomalaiset rikosentorjuntahankkeet ovat tyypillisesti pieniä pilottiprojekteja, joissa osallistujia on vain kourallinen. Tästä seuraa eräänlainen ”catch-22”-paradoksi: pienen pilottihankkeen skaalaaminen valtakunnalliseen mittakaavaan edellyttäisi vahvaan tutkimusasetelmaan perustuvaa tietoa ohjelman vaikuttavuudesta, mutta tietoa vaikuttavuudesta ei voi syntyä, jos osallistujamäärä on pieni ja vaikutukset tyypillistä kokoluokkaa. Siksi vaikuttavuustiedon kumuloitumisen näkökulmasta on vaikea perustella rahoituksen jakamista suurelle määrälle pieniä projekteja, joiden toimivuudesta ei voida saada luotettavaa tietoa. Tämä ei tietysti tarkoita sitä, etteikö ”ruohonjuuritason” voisi olla keskushallintoa kirkkaampia näkemyksiä vaikuttavista toimenpiteistä – mikäli vaikutusten tutkiminen olisi yhtään

helpompaa, voisi olla pikemminkin suotavaa antaa nykyistä enemmän tilaa pienille kokeiluille.

Mikäli satunnaistettuja kontrolloituja tutkimuksia ei jatkossakaan voida vankeinhoitossa toteuttaa, kysymys kuuluu, kannattaako esimerkiksi kaltaistusmenetelmiin perustuvien tutkimusten tekemistä vielä jatkaa? Jos vaihtoehtona on se, että ohjelmiin osallistuneiden vankeuden jälkeisiä lopputulemia ei jatkossa seurattaisi ollenkaan, vastaus on kyllä. Ja jos näiden ryhmien uusintarikollisuutta tutkitaan, lienee kuitenkin järkevämpää rajata verrokkiryhmää ilmeisten taustamuuttujien (ikä, sukupuoli, vankeuskertaisuus, tuomion pituus) osalta vertailukelpoiseksi kuin verrata keskimääräiseen uusintarikollisuuteen. Lisäksi on viisasta pohtia rajoituksia sen suhteen, millaisia ohjelmamuotoja verrokkit ovat voineet suorittaa. Vankeusajan sisällön ja ohjelmiin osallistumisen parempi mittaaminen voi auttaa näiden asetelmien vahvistamisessa ja vertailuasetelman selkeyttämisessä, mutta ohjelmiin valikoitumiseen liittyvää perustavanlaatuaista kysymystä paremmalla mittaamisella ei voi ratkaista. Siksi vahvempien tutkimusasetelmien käyttöä tulee jatkossa edistää sekä rikosseuraamusalalla että rikosentorjuntatoimenpiteiden arvioinnissa laajemminkin.

KIRJALLISUUS

- Aaltonen, M. & Tyni, S. (2017). Vaikutavuuden arviointi rikosseuraamusalalla. *Haaste*, 17(1), 14-17.
- Andrews, D.A., Zinger, I., Hoge, R.D., Bonta, J., Gendreau, P., & Cullen, F.T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, 28, 369-404.
- Angrist, J. & Pischke, J.-S. (2014). *Mastering 'Metrics: The Paths from Cause to Effect*. Princeton: Princeton University Press.
- Austin, P.C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res.*, 46(3): 399-424.
<https://doi.org/10.1080/00273171.2011.568786>
- Berman, A. (2004). The Reasoning and Rehabilitation program: Assessing short- and long-term outcomes among male Swedish prisoners. *Journal of Offender Rehabilitation*, 40, 85-103.
- Bushway, S.D. & Apel, R. (2012). A Signaling Perspective on Employment-Based Reentry Programming. Training Completion as a Desistance Signal. *Criminology & Public Policy*, 11(1), 21-50.
- Cann, J., Falshaw, L., Nugent, F., & Friendship, C. (2003). Understanding What Works: accredited cognitive skills programmes for adult men and young Offenders. *Home Office Findings*, 226.
- Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & De Rosario, H. (2020). *Pwr version 1.3-0. Basic Functions for Power Analysis*.
<https://github.com/heliosdrm/pwr>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, Second Edition*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Diamond, A. & Sekhon, J. (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.
<http://www.jstor.org/stable/43554804>
- Friendship, C., Mann, R., & Beech, A. (2003). Evaluation of a National Prison-based Treatment Program for Sexual Offenders in England and Wales. *Journal of Interpersonal Violence* 18, 744-759.
- Gelman, A., Skardhamar, T., & Aaltonen, M. (2017). Type M error might explain Weisburd's Paradox. *Journal of Quantitative Criminology*, 36, 295-304.
- Henning, K.R. & Frueh, B.C. (1996). Cognitive-behavioural treatment of incarcerated offenders: An evaluation of the Vermont department of corrections cognitive self-change program. *Criminal Justice and Behaviour*, 23, 523-541.
- Honka, O., Keinänen, A., & Tyni, S. (2006). Vaikuttaako vankeinhoito? *Haaste-lehti* 1/2006, 7-9.
- Hyatt, J.M. & Andersen, S.N. (2019). On the potential of incorporating administrative register data into randomized experiments. *Journal of Experimental Criminology*, 15, 469-497.
<https://doi.org/10.1007/s11292-019-09354-5>
- Knuuti, U. & Vogt-Airaksinen, T. (2013). *Ohjelmatyö rikosseuraamusalalla*. Rikosseuraamuslaitos.
- Konttila, A. & Tyni, S. (2011). *OMA-ohjelman suorittaneiden väkivaltarikollisten uusintarikollisuus sekä ohjelmanvälittömät vaikutukset* (Rikosseuraamuslaitoksen monisteita 4/2011). Rikosseuraamuslaitos.
- Konttila, A., Aaltonen, M., & Tyni, S. (2018). Väkivaltarikollisten OMA-ohjelman tehokkuuden ja vaikuttavuuden arviointi uusintarikollisuuden näkökulmasta: Analyysi OMA-ohjelman suorittaneiden ja keskeyttäneiden väkivaltarikollisten uusintarikollisuudesta. *Edilex*, 18577, 1-23.
- Koski-Jännes, A. (1995). *Huppu silmiltä. Katsaus vankien päihteiden käytön vähentämismenetelmiin ja niiden tuloksellisuuteen* (Oikeusministeriön vankeinhoito-osaston julkaisuja 1/1995). Oikeusministeriö.
- Laaksonen, T. & Tyni, S. (2015). Kognitiivis-behavioraalisen STOP-kuntoutusohjelman vaikuttavuudesta suomalaisilla seksuaalirikosvangeilla. *Sosiaalilääketieteellinen aikakauslehti*, 52(4), 306-316.
- Lavikkala, R. (2011). Vaikuttavan rikosseuraamustyön kysymyksiä. Teoksessa R. Lavikkala & H. Linderborg (toim.), *Rikosseuraamustyön kehittämisen kysymyksiä* (s. 89-90). Vantaa: Rikosseuraamusalan koulutuskeskus.
- Martinson, R. (1974). What works. Questions and answers about prison reform. *The Public Interest*, 10, 22-54.
- Martinson, R. (1979). Symposium on sentencing, Part II. New findings, new views: A note of caution regarding sentencing reform. *Hofstra Law Review*, 7(2).
- Menard, L. (2018). *Vankilateatterin vaikutukset ja vaikuttavuus*. *Taittuu ry:n esteettis-eettinen työtapo vankien kuntoutuksessa* (Rikosseuraamuslaitoksen julkaisuja 2/2018). Rikosseuraamuslaitos.
- Mews, A., Di Bella, L., & Purver, M. (2017). *Impact evaluation of the prison-based core sex offender treatment programme* (Ministry of Justice Analytical Series). London, England.
- Mitchell, J. & Palmer, E. (2004). Evaluating the "Reasoning and Rehabilitation" Program for Young Offenders. *Journal of Offender Rehabilitation*, 39(4), 31-45.
- Morgan, S.L. & Winship, C. (2014). *Counterfactuals and Causal Inference. Methods and Principles for Social Research*. Cambridge: Cambridge University Press.

- Nokso-Koivisto, O., Sarvimäki, M., & Toivanen, O. (2019). *Vaikutusarvioinnit osaksi päätöksentekoa – Miten varmistamme politiikassa, että teemme oikeita asioita*. Helsinki: Teollisuuden palkansaajat TP ry.
- Pitkänen, T., Kaskela, T., Tyni, S., & Tourunen, J. (2016). *Päihdehoitoon hakeutuneiden rikollisuus. Rekisteri- ja hoitotietoihin perustuva seurantatutkimus vankeusrangaistukseen tuomituista päihdehoidon asiakkaista* (Rikosseuraamuslaitoksen julkaisu 1/2016). Rikosseuraamuslaitos.
- Rikosseuraamuslaitos A. (2011–2020). *Rikosseuraamuslaitoksen tilastolliset vuosikirjat ajalta 2011–2020*. Helsinki: Rikosseuraamuslaitos (verkkojulkaisu). <https://www.rikosseuraamus.fi/fi/index/ajankohtaista/julkaisu/tilastollinen-vuosikirja.html> (Viitattu 1.10.2020).
- Rikosseuraamuslaitos B. (2020). *Rikosseuraamuslaitoksen tilinpäätös ja toimintakertomus vuodelta 2019* (Rikosseuraamuslaitoksen monisteita 1/2020). Rikosseuraamuslaitos.
- Ross, R., Fabiano, E., & Ewles, C. (1988). Reasoning and rehabilitation. *International Journal of Offender Therapy and Comparative Criminology*, 32, 29–35.
- Sirén, R. (2012). *Ehdoton vankeus, yhdyskuntapalvelu ja rikosuusiminen. Ehdoton vankeus ja yhdyskuntapalvelu törkeän rattijuopumuksen seuraamuksina* (Oikeuspoliittisen tutkimuslaitoksen tutkimustiedonantoja 116). Helsinki.
- Tyni, S. & Blomster, P. (2012). Vähentävätkö vankitoiminnot uusintarikollisuutta? *Yhteiskuntapolitiikka*, 77, 621–639.
- Tyni, S., Keinänen, A., & Kipinoinen, K. (2014). Cognitive Skills -ohjelman vaikutukset vankien ajatteluun, motivaatioon ja uusintarikollisuuteen: tulokset, ongelmat ja mahdollisuudet. *Edilex-artikkelit* 10.6.2014.
- Tyni, S. (2015). *Vankeinhoidon vaikuttavuus. Onko kuntoutukselle tilastollisia perusteita?* (Rikosseuraamuslaitoksen julkaisu 1/2015). Rikosseuraamuslaitos.
- Tyni, S. (2020). Vankien uusintarikollisuus vuosina 2000–2012 ja vankeuslain kokonaisuudistuksen uusintarikollisuusvaikutukset. *Oikeus*, 49(1), 47–73.
- Virta, L. (2013). *KUVA - Selville vesille Kuntouttava vankityö -ohjelman arviointi* (Rikosseuraamuslaitoksen monisteita 5/2013). Rikosseuraamuslaitos.
- Virtanen, M., Aaltonen, M., Tyni, S., & Kivivuori, J. (2020). Vankilakoulutuksen vaikutus uusintarikollisuuteen ja vapautumisen jälkeiseen työssäkäyntiin. *Edilex* 2020/3.
- Väyrynen, S., Ronkainen, S., Silen, M., Keränen, A., & Tyni, S. (2014). *Muutokseen motivointi ja väkivaltaisuuksien muutos. Arviointitutkimus Move! -ohjelman vaikuttavuudesta* (Rikosseuraamuslaitoksen julkaisu 2/2014). Rikosseuraamuslaitos.
- Wilkinson, J. (2005). Evaluating Evidence for the Effectiveness of the Reasoning and Rehabilitation Programme. *The Howard Journal of Criminal Justice*, 44, 70–85.

ENGLISH SUMMARY

Sasu Tyni and Mikko Aaltonen: Evaluation research on rehabilitation programs in prison.

Rehabilitation programs have been used systematically in Finnish prisons now for two decades. The main aim of these programs is to promote desistance and reduce recidivism after release. Several Finnish evaluations of prison programs have been published during the recent years. In this article, we start by reviewing the key results of these studies and consider the possibilities and limitations of quasi-experimental program evaluation in the light of available register data and the scale of program uptake. So far none of the Finnish evaluations have shown evidence of programs reducing recidivism. Even though the datasets needed for evaluation have improved clearly in the 21st century, credible evaluation of programs' causal effects is still hampered by lack of strong research designs and low participant rates in most programs. The methodological quality of evaluations should be improved. Given that the available evidence suggests that true effect sizes of programs are likely to be relatively small, large samples are needed to attain enough statistical power to detect effects or their absence. At the same time, even if true effect sizes are small, the programs can still be cost-effective. The low numbers of participants in most prison programs present a challenge to evidence-based program development.

Keywords: prison rehabilitation – recidivism – program evaluation – statistical power – causal inference