

# Virker ART i svenske fengsler?

Af Frode Svartdal<sup>1</sup> & Knut Gundersen<sup>2</sup>

## Abstract

*This paper discusses a report on the effect of Aggression Replacement Training (ART) in Swedish prisons (Danielsson, Fors, & Freij, 2011). The report compared two samples: An »ART group« that received ART in the period 2003-2006, and a »control group« that was selected to match the ART group at a later time from the general Swedish prison population. A close examination of the report reveals that it failed to specify the nature and quality of the treatment given to the ART group. Attrition from the ART group was very high, indicating that the quality of the intervention was low. The report reveals that the two groups were very different, with significantly higher level of violence and aggression history in the ART group compared to the control group. We conclude that this difference is a likely cause of the higher rate of violent recidivism in the ART group, and that the ART intervention most probably had nothing to do with the observed differences between the groups. This case illustrates the risks of making important policy decisions based on findings from reports published outside the peer review system of established journals.*

## 1. Innledning

Forskningsresultater offentliggjøres gjerne i publikasjonskanaler med fagfelle-vurdering (*peer review*). En viktig grunn til dette er kvalitetssikring: Ved at fagfeller kritisk gjennomgår metode, resultater og tolkningen av resultatene, kan vi være nokså sikre på at de konklusjoner som publiseres, er holdbare. Noen forskningsresultater publiseres likevel ofte utenom disse kvalitetssikrede kanalene. Dette gjelder i særlig grad resultater fra anvendt forskning. Her ser vi mange eksempler på at forskning (gjerne under alternative navn, så som »rapport«, »utredning« eller »evaluering«) bekjentgjøres i interne rapporter, bokkapitler og bøker uten tilfredsstillende fagfelle-vurdering.

En slik praksis er forståelig fordi praktisk virksomhet – og forskning fokusert på slik virksomhet – er uhyre kompleks. Men samtidig er det et paradoks at denne kompleksiteten medfører reduserte krav til kvalitetssikring. En slik praksis inne-

\* Title in English: *Does Aggression Replacement Training (ART) work in Swedish prisons?*

bærer at praktiske tiltak vi ønsker skal være forskningsbaserte, i alle fall i noen tilfeller kan være basert på forskning som ikke holder mål. Når forskningsresultater bekjentgjøres uten slik kvalitetssikring, er det derfor grunn til å være på vakt.

Nylig ble rapporten **Behandlingsprogrammet ART i kriminalvården. Utvurdering av återfall i brott för programdeltagare 2003-2006** (Danielsson, Fors & Freij, 2011) publisert. Dette er en rapport som ikke har gjennomgått den uavhengige fagfellevurdering som rapporten ville måttet gå gjennom dersom den var blitt sendt inn for publikasjon. Fordi rapporten presenterer konklusjoner som dels viser andre resultater enn tidligere undersøkelser om ART i fengsel (Barnoski & Aos, 2004; Hatcher et al., 2008) samtidig som den fikk avgjørende betydning for at ART ble fjernet fra svenske fengsler, kan dette være et eksempel som illustrerer aktualiteten i omtalte problemstilling.

Rapporten presenterer en evaluering av effekten av ART (*Aggression Replacement Training*) brukt i svenske fengsler i perioden 2003-2006. Over 3000 deltakere inngikk i undersøkelsen. Hensikten med rapporten fremgår på s. 6:

»Syftet med denna registerstudie var att i en kvasiexperimentell design undersöka om klienters återfallsrisk minskar för dem som genomgått vuxenversionen av behandlingsprogrammet Aggression Replacement Training (ART) inom svensk kriminalvård.«

Resultatene, analysert ved hjelp av Cox regression, var negative (s. 6):

»Resultatet visade att behandlingsgruppen hade en statistiskt säkerställd ökad risk att återfalla i något brott i jämförelse med jämförelsegruppen (Hazard kvot [HK]=1,25; 95 % KI: 1,07-1,46) och att risken ökade ytterligare i en jämförelse mellan enbart de klienter som avbrutit programmet i förtid och jämförelsegruppen (HK=1,43; 95 % KI: 1,17-1,76).«

I denne artikkelen vurderer vi de resultater og konklusjoner Danielsson, Fors og Freij (2011) har presentert. I vurderingen tar vi utgangspunkt i tre spørsmål som effektevalueringer bør adressere (Barnoski et al., 2004):

- Var det faktisk ART som er gjennomført og evaluert?
- Var programmet implementert på en kvalitetssikret måte?
- Er evalueringen gjennomført ut fra de kriterier man forventer av god forskning?

Vi drøfter disse spørsmålene i det følgende.

## **2. Var det ART som ble gjennomført og evaluert?**

En vurdering av om det faktisk var ART som ble evaluert må ta utgangspunkt i informasjon om tre viktige faktorer: Dose, innhold og kvalitet.

### *Dose*

Et standard ART-program har 30 leksjoner, hver på en time (Goldstein, Glick, & Gibbs, 1998). Den foreliggende undersøkelse omfattet 30 leksjoner, hver på tre timer. De færreste gjennomførte imidlertid 30 leksjoner (gruppemøter). De som ble klassifisert som »fullførte«, 59 % av deltakerne, hadde i gjennomsnitt deltatt på 24 gruppemøter, mens de resterende 41 % (de som »avbrøt«) hadde i 8 gruppemøter i gjennomsnitt. Hvorvidt det er rimelig å klassifisere deltakere som »fullførte« når de ikke hadde deltatt på 1/5 av de foreskrevne møtene, kan diskuteres. Men at innsatte som hadde deltatt i bare 8 av i alt 30 av intervensjonens møter i det hele tatt kan inkluderes i effektundersøkelsen er sannsynligvis lite meningsfullt.

Nå kan man hevde at redusert deltakelse på gruppemøtene i dette tilfellet kompenseres ved at hvert møte hadde en lengde på tre timer (i stedet for en time, som er standard). Imidlertid bør avviket fra den anbefalte lengden på hvert møte problematiseres. Det er ikke åpenbart at økt lengde på hver sesjon kompenserer for reduksjon i antall møter, bl.a. fordi det kan være vanskeligere å opprettholde engasjement og motivasjon over lengre sesjoner enn ved korte.

### *Innhold*

Forskerne oppga at det ble gjennomført et standard ART-program med innslag av motiverende intervju i tillegg. Ut over dette presenterer rapporten ingen informasjon om innholdet i intervensjonen. Vi vet for eksempel ikke om alle de tre komponentene som inngår i standard ART – sinnekontroll, ferdighetstrening og moralsk resonering – inngikk. Vi vet heller ikke noe om andelen av spesifikke ART-øvelser i forhold til annet innhold (motiverende intervju).

### *Implementeringskvalitet*

Rapporten sier ingenting om hvilken opplæring instruktørene har hatt før de startet programmet, hvordan deltakerne evaluerte programmet eller om det var lagt inn kvalitetssikringssystemer for å sikre at programmets innhold ble fulgt og at innholdet ble tilpasset den aktuelle målgruppe. Forskning viser at måten et program implementeres og kvalitetssikres på, sannsynligvis har like stor betydning som programmet per se (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005).

På spørsmål om hvordan ART-intervensjonene ble implementert, har en av forfatterne av rapporten, Freij, svart slik (Freij, 2011):

»Därför vill jag poängtera att det är ART såsom programmet bedrevs inom Kriminalvården 2003-2006 som är utvärderat i denna rapport. Vi har inte information om hur väl varje enskild programledare har följt manualen, hur man har inkluderat klienter till programmet, om man har genomfört samtliga tre delar, om man har individanpassat programmet osv. Det vi vet är hur utbildningen och handledningen för programledarna såg ut under den här perioden och vilka instruktioner/manualer personal hade att gå efter. Jag vill tydliggöra att det är programledarna själva som har gjort bedömningen om en klienter har fullföljt eller avbrutit programmet, inte vi som utvärderare, och att 24 gruppmöten var det genomsnittliga antalet över samtliga klienter som bedömdes ha fullföljt.«

Danielsson, Fors og Freij (2011) har altså brukt data fra deltakere i intervensjonsprogrammet uten å vite om disse faktisk har mottatt ART, og i tilfelle av hvilken kvalitet. Når informasjon om intervensjonen ikke er kjent, er det også vanskelig å si noe om hvordan resultatene av tiltaket kan forklares. I diskusjonen reflekterer forfatterne selv over dette:

»Det bör poängteras att utvärderingen är gjord på registerdata från åren 2003 till och med 2006 och man kan anta att arbetet med att implementera och utveckla programmets olika delar gått framåt under de senaste åren. Kanske skulle en utvärdering av åren 2007-2009 visa på andra resultat.« (s. 27).

Vi vet i dag at implementeringsaspektet er svært viktig for å kunne vurdere effektene av et tiltak, og vi vet også at tiltak ofte – som også tiltaket diskutert i denne artikkelen illustrerer – gjennomføres uten tilstrekkelig oppmerksomhet rundt implementeringskvalitet (Sørli, Ogden, Solholm, & Røyhus Olseth, 2010; Wallace, Blase, Fixsen, & Naom, 2008). For å sikre at programmet leveres med den dosering, kvalitet og innhold som er foreskrevet, må også organisasjonen som leverer programmet ha innført nødvendige støttefunksjoner. Dette inkluderer tid for planlegging av timer, veiledningsrutiner, mulighet for å generalisere ferdigheter utenfor ART-rommet, samt aktuelle belønningssystemer. En forutsetning for at disse momentene skal kunne oppfylles, er at leder kjenner til programmet og sikrer at kvalitetssystemene er på plass og at den øvrige stab støtter programmet og bidrar til at nye ferdigheter blir generalisert (Gundersen, 2005; Goldstein et al., 2000).

Man kan dermed tenke seg at innhold og kvalitet i et program kan være tilfredsstillende, men at det omliggende system ikke støtter programmet og således bidrar til at den nye læring faktisk ikke blir praktisert. I Barnoski og Aos sin un-

dersøkelse av forskningsbaserte program for ungdom i fengsel (Barnoski et al., 2004) skilte forskerne mellom »competent delivery« og »non-competent delivery« ut fra en evaluering ut fra de kvalitetssikringssystemer som var bygget opp rundt intervensjonen. Baronoski et al. estimerer i sin undersøkelse en reduksjon i tilbakefall til fengselsstraff på 24 % etter 18 måneder sammenliknet med kontrollgruppen, forutsatt at programmet var implementert slik det skal. Dersom implementeringen **ikke** var tilfredsstillende, var tendensen svak forverring sammenliknet med kontrollgruppen.

Vi må anta at Danielson et al. (2011) er kjent med forskningen vedr. implementeringskvalitet. Likevel synes ikke dette å influere på hovedkonklusjonen som presenteres i rapporten.

### **3. Ble evalueringen gjennomført ut fra de kriterier man forventer av god forskning?**

Hvorvidt undersøkelsen oppfyller de kriterier man forventer fra god forskning, er et komplisert spørsmål å besvare. Her må man vurdere design, variabler og operasjonalisering av disse, analyse av resultater og tolkning av disse, med mer. Vi drøfter noen sentrale spørsmål her.

#### *Design*

For at en undersøkelse skal gi informasjon om effekt av en intervensjon, i dette tilfellet effekt av ART, kreves sammenligning. Slik sammenligning er i gruppestudier gjerne mellom to tidspunkter (før vs. etter behandling), og/eller mellom grupper (ART-gruppe vs. kontrollgruppe) etter behandling. Danielsson et al. (2011) har valgt sistnevnte strategi. En slik gruppesammenligning etter intervensjon er mye brukt, eks. i »gullstandarden« i utprøving av medikamenter og intervensjoner, nemlig randomiserte gruppestudier (RCT).

I denne undersøkelsen har man ikke kunnet randomisere (tilfeldig fordele) deltakerne til ART vs. behandlingsgrupper. Dette skyldes det enkle faktum at tiltaket for ART-gruppen ble gjennomført for mange år siden. Deltakerne er dermed inkludert i ART vs. kontrollgrupper på en ikke-tilfeldig måte. Danielsson et al. (2011) omtaler (s. 6) studien som en »en kvasiexperimentell design«. Normalt betyr dette at undersøkelsen skal gjennomføre en manipulasjon (ART vs. ikke-ART), men at den ikke har den grad av kontroll som en »ekte« eksperimentell studie har. I dette tilfellet er det vel også en overdrivelse å kalle studien »kvasiexperimentell«, i og med at forskerne selv ikke har gjennomført noen manipulasjon. Strengt tatt er dette derfor en korrelasjonell studie der forskerne har sammenliknet to grupper, en gruppe som gjennomgikk ART i perioden 2003-2006,

og en annen gruppe som ikke gjennomgikk ART men som ble utvalgt for å være sammenlignbar med ART-gruppen.

I seg selv er dette uproblematisk; det er gjennomført mye god forskning ved bruk av både korrelasjonelle og kvasieksperimentelle design. Men samtidig er det velkjent at konklusjoner vedr. effekt fra slike design er mye mer usikre enn om designet hadde vært et »ekte« eksperimentelt design. Grunnen er enkel: Hvis man ikke har kontroll på andre relevante påvirkningsfaktorer (confounding variabler) gjennom randomisering, må disse kontrolleres på andre måter for at en slutning om effekt skal være holdbar.

Holdbare slutninger om effekt av et tiltak er ofte problematisk. I kvasi-eksperimentelle design kan man inkludere en pre-test før intervensjonen samt en test etter intervensjonen, slik at man får informasjon om det faktisk har foregått en endring i tidsmessig sammenheng med intervensjonen. I dette tilfellet har vi ikke noen pre-test. Vi vet derfor ikke om det har inntruffet en endring i sammenheng med ART-tiltaket, om andre hendelser kan forklare den observerte gruppeforskjellen, eller om det simpelt hen er slik at gruppene har vært forskjellige hele tiden. Forskerne tilskriver likevel forskjellen i tilbakefallsrisiko mellom gruppene til intervensjonen uten å ta på alvor at denne tolkningen trues av en rekke alternative tolkninger. Som kjent finnes det en lang liste av velkjente feilkilder forskere må sjekke når den indre validiteten i en eksperimentell undersøkelse skal dokumenteres. Shaughnessy et al. (2009 s. 349) beskriver det problematiske med kvasi-eksperimentelle design på denne måten: »Researchers must recognize the specific shortcomings of quasi-experimental procedures, and they must work like detectives to provide whatever evidence they can to overcome these shortcomings.« I dette tilfellet ser vi få indikasjoner på at forskerne tar alternative tolkninger av den observerte gruppeforskjellen på alvor.

### *Rimeligheten i effektvurderingene*

Konklusjonen undersøkelsen formulerer om effekt av ART er dels basert på hele ART-utvalget, den delen av det som fullførte (59 %), og/eller den delen som ikke fullførte. Vurdering av »effekt« i disse to utvalgene spriker. Overflattisk betraktet viser resultatene at det er forbundet med en dramatisk forhøyet risiko (43 %) for tilbakefall hvis man følger ART-treningen i 8 gruppemøter. Forfatterne drøfter ikke det åpenbart urimelige i et slikt resultat. Hvordan kan man tenke seg at det å følge programmet i 8 møter skulle ha en så dramatisk negativ effekt? Samtidig viser undersøkelsen at det å følge treningen i 24 møter *ikke* gir merkbar risikoøkning i forhold til kontrollgruppen. Vi må anta at de fleste vil spørre seg om det ikke kan tenkes at det er andre forhold som kan forklare disse resultatene. Et nær-

liggende forhold, som at utvalgene kan ha vært forskjellige i utgangspunktet, drøftes grundig i neste avsnitt.

### Seleksjon

En av de mest seriøse truslene mot slutninger om effekt av en intervensjon i undersøkelser som dette, er at gruppene har vært forskjellige hele tiden, uavhengig av intervensjonen. Det er derfor av største interesse å se nærmere på hvordan deltakerne til de to gruppene som inngikk i denne studien ble utvalgt.

**ART-gruppen** bestod av alle som hadde gjennomgått ART i årene 2003-2006 i svenske fengsler, samlet 523 personer. Her har forskerne ikke gjort noen ytterligere utvelgelse; alle som deltok i ART-intervensjonene, ble inkludert. **Kontrollgruppen** bestod av sammenlignbare personer innsatt i samme periode, samlet 2615 personer. Denne gruppen ble opprinnelig hentet fra et stort utvalg av potensielle deltakere, ca. 26000 (se rapporten s. 16). Her har forskerne gjennomført en relativt streng seleksjon for å matche kontrollgruppen med ART-gruppen. Rapporten oppgir en del inklusjons- og eksklusjonskriterier.

Det viktigste kriterium for deltakelse i kontrollgruppen var det selvsagte faktum at personene i perioden 2003-2006 **ikke** hadde blitt utvalgt til å gjennomgå ART. Dette gjør det naturlig å spørre hvilke av kriteriene som i sin tid lå til grunn for rekruttering til ART-gruppene. Blant de over 26000 som **kunne** fått ART i perioden 2003-2006, fikk altså 523 personer dette tilbudet og takket ja. Rapporten (s. 10; kursiv lagt til av oss) fremhever at denne gruppen ble sterkt selektert på grunnlag av aggresjons- og voldsproblematikk:

»Inom Kriminalvården används ART främst för klienter dömda för våldsbrott. En klient som *inte* dömts för våldsbrott kan inkluderas om han eller hon har uppvisat en aggressivitetsproblematik. Klienten ska då ha uppvisat ett mönster med våldsamt beteende och svårighet att behärska sig under en längre period och i olika situationer. Ett ytterligare inklusionskriterium är att klienten själv ska uppleva sin aggressivitet som ett problem och ha en vilja att förändra sitt beteende. Endast klienter som bedömts ha medelhög eller hög risk för återfall i våldsamt beteende bör gå i ART.«

Seleksjon til kontrollgruppen ble gjort ut fra krav om at deltakerne skulle være menn (siden alle i ART-gruppen var menn), og at de ikke hadde påbegynt andre behandlingsprogrammer, og skulle være innsatt i minst 12 uker. Rapporten (s. 16-17) skriver så:

»Jämförelsegruppen bestod ursprungligen av cirka 26 000 personer som uppfyllde ovanstående kriterier. Ur denna grupp gjordes ett stratifierat urval på indexbrott (våld/ickevåld) och narkoti-

kamissbruk (uppgift saknas/ingen missbruksproblematik /inget gravt missbruk/gravt missbruk) för att matcha behandlingsgruppens andel personer med dessa karaktäristika inom varje grupp.«

Fra dette kan man trekke følgende viktige konklusjon: ART-gruppen ble i sin tid sterkt selektert ut fra to kriterier på vold, (a) dom for voldshandling, eller (b) i fravær av voldsdom, dokumentert voldsproblematikk over lang tid og i ulike situasjoner. Hele ART-utvalget har altså hatt en dokumentert historie med aggresjon og vold, og alle inkludert i ART-behandlingene har blitt bedømt å ha »medelhög eller hög risk för återfall i våldsamt beteende«.

Kontrollgruppen har gjennomgått en mildere seleksjon. Spesielt interessant er det at kontrollgruppen ble etablert gjennom et stratifisert utvalg med samme andel voldsdomte som i ART-gruppen. Dette betyr i praksis at selv om andelen voldsdomte i ART-gruppen er den samme som i kontrollgruppen, er den andelen som *ikke* er voldsdomt i ART-gruppen preget av langvarig og generell volds- og aggresjonsproblematikk, mens tilsvarende andelen i kontrollgruppen er definert ved at de *ikke* har en voldsdom. Denne andelen utgjør ca. halvparten av utvalgene (se Tabell 2.1b, s. 33). Konkret betyr dette at denne halvparten av utvalgene er forskjellige: *Hele* ART-gruppen er preget av volds- og aggresjonsproblematikk, men *halvparten* av kontrollgruppen er definert ved at de *ikke* har noe voldsproblem.

Dette viser altså at ART og kontrollgruppene var systematisk forskjellige relatert til vold og aggresjon, uavhengig av intervensjonen. Undersøkelsen inkluderer få data som måler hvordan den skjeve seleksjonen gir seg utslag i faktiske gruppeforskjeller. Men Tabell 2:1a viser at de to gruppene i alle fall på to mål var signifikant forskjellige. ART-gruppen var signifikant yngre enn kontrollgruppen (25,4 vs. 34,6 år), og den hadde signifikant lengre tid i anstalt (376 vs. 276,9 dager).

Fra Tabell 2:1a.

Undersøkt faktor	Behandlings- vs. jämförelsegrupp	
	Behandling (n=523) Medelvärde	Jämförelse (n=2615) Medelvärde
Ålder (år)	25,4	34,6 **
Tid i anstalt (dagar) <sup>a</sup>	376,0	276,9 **
Tidigare verkställigheter (antal) <sup>b</sup>	1,2	1,0



Hvis forskerne hadde gjennomført en god seleksjon til kontrollgruppen, skulle man ideelt sett forvente at denne var relativt lik ART-gruppen på alle relevante variabler i tillegg til de to variablene som ble brukt i matchingen, lovbrudd (vold/ikke-vold) og narkotikamisbruk. Dette er åpenbart ikke tilfelle. Aldersforskjellen var dramatisk (nesten 10 år), og antall dager i anstalt var mye høyere i ART-gruppen. Sistnevnte forskjell er spesielt interessant fordi den viser at antall dager sonet er mye høyere i ART-gruppen, til tross for at denne gruppen også er mye yngre enn kontrollgruppen. Dette støtter en antakelse om at ART-gruppen har vært en gruppe med spesielt tung volds- og aggresjonsproblematikk.<sup>3</sup> Dette støttes også av det faktum at prosentandelen av innsatte som hadde begått ran var signifikant høyere i ART-gruppen enn i kontrollgruppen (30,8 vs. 17,2). Samlet viser dette at vi har å gjøre med to forskjellige utvalg, der ART-gruppen var sterkt selektert i retning av høy grad av volds- og aggresjonsproblematikk.

*Aldersforskjellen mellom utvalgene skyldes at ART-gruppen er spesiell*

Som påpekt over var det mer enn 9 års forskjell i gjennomsnittsalder mellom utvalgene. Spørsmålet er: Hvordan kunne en så dramatisk forskjell oppstå? *ART-utvalget* bestod av 523 personer, med en gjennomsnittlig alder = 25,4 år. *Kontrollgruppen*, 2615 personer, hadde et aldersgjennomsnitt på 34,6 år. Begge gruppene ble hentet fra det som definerer populasjonen for denne undersøkelsen, ca. 26000 innsatte i svenske fengsler i perioden 2003-2006. Hadde deltakerne blitt tilfeldig utvalgt, skulle aldersgjennomsnittene være ganske like og samsvarende med populasjonsgjennomsnittet.

Informasjon om gjennomsnittsalderen for hele fengselspopulasjonen kunne gi indikasjon på hvilken av gruppene som er mest representativ for populasjonen, og dermed hvilken av gruppene som avviker fra denne. Med så store utvalg må det ha vært sterke seleksjonsmekanismer i operasjon for å få en så stor aldersforskjell mellom gruppene. Hvis det skulle vise seg at ART-gruppen er den som avviker mest, er det av den største interesse å få belyst seleksjonsmekanismene som har vært i funksjon ved utvelgelse av denne gruppen. Kunnskap om slike seleksjonsmekanismer ville kunne indikere variabler av betydning for å forklare forskjell i tilbakefallsrisiko mellom gruppene.

En kilde anslår gjennomsnittsalderen til innsatte i svenske fengsler til 36 år.<sup>4</sup> Selv om dette anslaget antakelig gjelder for årene før 2003 og inkluderer både kvinner og menn, samsvarer dette rimelig godt med kontrollgruppens gjennomsnittsalder i denne undersøkelsen. Dette betyr at det er ART-gruppen som skiller seg ut aldersmessig; kontrollgruppen ligner fengselspopulasjonen.

*Statistiske analyser: Garbage in, garbage out*

De foregående punktene har godtgjort at ART-gruppen var spesiell og sterkt avvikende fra den generelle fengselspopulasjonen ved at den (a) var selektert i retning av tung volds- og aggresjonsproblematikk, og (b) ART-gruppen var også markant yngre enn kontrollgruppen. Begge forhold utgjør mulige forklaringer på hvorfor tilbakefall i form av dom/voldsdom var hyppigere i ART-gruppen enn i kontrollgruppen. Eksempelvis viser forskning (eks. Gottfredson & Hirschi, 1990) at den aldersforskjellen som eksisterer mellom utvalgene i seg selv kan være en mulig forklaring på forskjellen i tilbakefallsrisiko mellom gruppene.

Slike alternative forklaringer i kvasiexperimentelle undersøkelser viser til confounding variabler, påvirkningsfaktorer som alternativt kan forklare en observert effekt. Rapporten (s. 18) understreker dette poenget slik:

»Det är helt nödvändigt att ta hänsyn till faktorer som ålder, kön, missbruk, omfattning och typ av tidigare brottslighet etc. i analyserna då dessa är kopplade till återfallrisk. Annars riskerar man att tolka skillnader i återfall mellan behandlings- och jämförelsegrupp som en konsekvens av behandlingen, när de snarare beror på faktorer som fanns där redan innan behandling. Detta gäller framförallt i studier av klienter som inte fördelats slumpmässigt (randomiserats) till behandlings- och jämförelsegrupp.«

For å løse disse problemene, anvender forskerne en statistisk prosedyre som tar hensyn til faktorer som alternativt kan forklare den observerte gruppeforskjellen (s. 18):

»För de statistiska analyserna i denna undersökning användes därför *Cox regression*, en analysmetod som tar hänsyn till andra påverkansfaktorer (på engelska kallade »confounding« faktorer eller »confounders«). Kännetecknande för en sådan faktor är att den samvarierar eller korrelerar (positivt eller negativt) med både den beroende variabeln (dvs. det man vill undersöka förekomst av – här återfall i brott) och den oberoende variabeln (dvs. det man vill testa effekten av – här behandling).«

Rapporten viser (s. 32-33) at tre kontinuerlige og 6 diskrete variabler (en med flere underkategorier) er anvendt i regresjonsanalysen for å »ta hensyn til andre påvirkningsfaktorer« i konklusjon om effekt av ART. Spørsmålet er om disse faktorene gir mening som forklaringsvariabler, og om de fanger opp den allerede eksisterende skjevfordelingen mellom gruppene. De tre kontinuerlige variablene som er angitt, er alder, tid i anstalt, og antall tidligere soninger (Tabell 2:1a):

Undersøkt faktor
Ålder (år)
Tid i anstalt (dagar) <sup>a</sup>
Tidigare verkställigheter (antal) <sup>b</sup>

Følgende diskrete variabler er angitt (Tabell 2:1b):

Undersøkt faktor
Nordiskt medborgarskap
Utvising i dom
Påföljd: Fängelse
Påföljd: Skyddstillsyn
Påföljd: Villkorlig frigivning med övervakning
Tidigare brott <sup>a</sup>
Våldsbrott
Sexualbrott
Rån, även grovt
Tillgrepsbrott
Övriga förmögensbrott
Brott mot allmänhet och stat
Narkotikabrott/varusmuggling
Rattfylleri, även grovt
Trafikbrott, ej rattfylleri
Övriga brott

Man kan alltid diskutere hvilke av disse faktorene som kan tenkes å være mulige påvirkningsfaktorer i forhold til sannsynlighet for gjentatt lovbrudd, men ut fra forskningsmessige betraktninger er i alle fall to av disse aktuelle:

- *Alder* er aktuell fordi forskning har vist at alder henger sammen med kriminalitet (Gottfredson & Hirschi, 1990). Som kjent var ART-gruppen i gjennomsnitt signifikant yngre enn kontrollgruppen.
- *Tid i fengsel* kan også være aktuell. Det var ikke noen forskjell mellom gruppene hva angår antall tidligere dommer, men en klar forskjell i tid i anstalt, noe som indikerer at dette målet reflekterer alvorlighetsgrad i de lovbrudd man er dømt for. Tid i anstalt var signifikant høyere hos ART-gruppen, noe som indikerer at disse har fått strengere dommer. Tabell 2:1b viser at disse strenge dommene godt kan henge sammen med tilbøyelighet til vold og aggresjon; eksempelvis var det en signifikant større andel av ART-utvalget som var blitt dømt for ran.

Flere av de andre variablene som er listet opp som mulige »påvirkningsfaktorer« kan i og for seg være interessante, men de er høyst sannsynlig uinteressante som mulige påvirkningsfaktorer for å forklare tilbakefall. Det mest slående med listen er *fraværet* av interessante faktorer. Hvis denne undersøkelsen hadde vært utført som en ordinær kvasiekperimentell undersøkelse med kontroll for mulige confounding variabler, ville man innhentet mål fra variabler som antas – ut fra forskning og teori – å kunne spille en rolle for om ART virker eller ikke. I tillegg til alder (som er med) kunne man tenke seg en rekke interessante variabler relatert til selvkontroll og impulsivitet, personlighet, motivasjon, utdanningsnivå, venne- og familierelasjoner (se Pratt & Cullen, 2000). Ingen slike foreligger.<sup>5</sup>

#### 4. Konklusjon

De diskuterte momentene peker mot den konklusjon at denne undersøkelsen ikke er holdbar: Vi vet ikke hvilken intervensjon som ble innført under navnet ART, og vi vet ikke hvilken kvalitet den hadde. Det vi vet, er at det var en rekke til dels dramatiske forskjeller mellom de to gruppene som sammenlignes i denne undersøkelsen, ART-gruppen og kontrollgruppen. Forskerne dokumenterer selv mange av disse forskjellene, men hevder at de »kontrolleres for« i avanserte statistiske analyser, slik at man kan få en enkelt indeks som uttrykker effekt av intervensjonen sammenlignet med kontrollgruppen. Vi hevder at de variablene analysene trenger for å kontrollere for de dramatiske gruppeforskjellene er fraværende. Bortsett fra variabelen alder, er de andre variablene i liten eller ingen grad potensielle forklaringsvariabler. Mest sannsynlig kan forskjellene vi ser mellom gruppene tilskrives det faktum at ART-gruppen i sin tid ble sterkt selektert på dokumentert tilbøyelighet for vold og aggresjon, mens kontrollgruppen ligner på den generelle fengelsespopulasjonen. I tillegg må ART-gruppen være sterkt selektert mht alder, i og med at den avviker dramatisk ved å være mye yngre enn den generelle fengelsespopulasjonen. Å sammenligne to grupper som er så forskjellige av andre grunner enn ART-intervensjonen, gjør det meget vanskelig å si noe meningsfullt om eventuelle effekter av intervensjonen (hva den nå egentlig bestod i). En fagfellevurdering i et vitenskapelig tidsskrift ville etter all sannsynlighet ha påpekt de forskningsmessige svakheter i denne rapporten og dermed gitt de besluttende myndigheter et sikrere grunnlag for å gjøre riktige beslutninger. Forskere har stor makt, både i forhold til å anbefale programmer og å forkaste dem. En fagfellevurdering vil være en form for kvalitetssikring for å hindre feil, enten ved at gode programmer blir forkastet eller dårlige programmet blir opprettholdt.

## Noter

1. Professor Dr. Philos., Universitetet i Tromsø, Norge
2. Førstelektor, Diakonhjemmet Høgskole, Norge
3. Det fremgår av Tabell 2:1b at ART-gruppen har en signifikant høyere andel av soning utenfor anstalt («Vilkorlig frigivning med overvaking»). Dette kan tyde på at tid i anstalt i Tabell 2:1a, som allerede er høy i ART-gruppen, egentlig kunne vært enda høyere.
4. <http://www.dip-alicante.es/agis/documents/primero/suecia2gb.pdf>
5. Men man kan anta systematiske forskjeller mellom utvalgene her. I Tabell 2:1b fremgår det eksempelvis at ART- og kontrollgruppene var signifikant forskjellige ved at ART-gruppen hadde langt mindre andel deltakere med seksualdommer, men langt høyere andel som var dømt for ran.

## Referanser

- Barnoski, R. & Aos, S. (2004). *Outcome Evaluatoins of Washington States' Research-Based Programs for Juvenile Offenders* (Rep. No. 04-01-1201). Washington State Institute for Public Policy.
- Danielsson, M., Fors, A., & Freij,I (2011). *Behandlingsprogrammet ART i Kriminalvården. Utvärdering av återfall i brott för programdeltagare 2003-2006*. Kriminalvårdens Utvecklingsenhet.
- Fixsen, D., Naoom, S., Blase, K., Friedman, R.M., & Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature*. Tampa, Florida: USF, University of Florida.
- Freij, I. (2011). Mail til B. Daleflod, 15. februar.
- Goldstein, A.P. & Martens, B. (2000). *Lasting change: Methods for Enhancing Generalization of Gain*. Champaign, IL: Research Press.
- Gottfredson, M.R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.
- Goldstein, A.P., Glick, B., & Gibbs, J. C. (1998). *Aggression replacement training a comprehensive intervention for aggressive youth*. (Rev. ed ed.) Champaign, Ill: Research Press.
- Gundersen, K. (2005). Implementering av ART. I L. Moynahan, B. Strømgren & K. Gundersen (red.) *Erstatt aggresjonen, Aggression Replacement Training og positive atferds- og støtte-tiltak* (s. 140-153) Oslo: Universitetsforlaget.
- Hatcher, R.M., Palmer, E.J., McGuire, J., Hounsone, J.C., Bilby, C.A.L., & Hollin, C. (2008). Aggression replacement training with adult male offenders within community settings: a re-conviction analysis. *The Journal of Fornesic Psychiatry and Psychology*, 19, 517-532.
- Pratt, T.T. & Cullen, F.T. (2000). The empirical status of the Gottfredson and Hirschi's General theory of crime: A meta-analysis. *Criminology*, 38, 931-964.
- Shaughnessy, J.J., Zechmeister, E.B., & Zechmeister, J.S. (2009). *Research methods in psychology (8th ed.)*. New York: McGraw Hill.
- Sørlie, A.M., Ogden, T., Solholm, R., & Røyhus Olseth, A. (2010). Implementeringskvalitet – om å få tiltak til å virke. *Tidsskrift for norsk psykologforening*, 47.
- Wallace, F., Blase, K., Fixsen, D., & Naoom, S. (2008). *Implementing the Findings of Research: Bridging the Gap Between Knowledge and Practice*. Alexandria, VA: Educational Research Service.