

Virker ART i svenske fengsler?

Frode Svartdal, professor dr.philos.
Universitetet i Tromsø

Nylig ble rapporten BEHANDLINGSPROGRAMMET ART I KRIMINALVÅRDEN. UTVÅRDERING AV ÅTERFALL I BROTT FÖR PROGRAMDELTAGARE 2003-2006 (Danielsson, Fors & Freij, 2011) publisert. 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 fra undersøkelsen er 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).”

Siden denne negative konklusjonen er uventet (den bryter med det overveiende positive bilde ART-evalueringer har gitt internasjonalt og i Norge), er det all grunn til å studere denne rapporten grundig. Dette ønsket forsterkes av det faktum at denne undersøkelsen har omfattet et svært stort utvalg, hele 3138 deltakere. Dessuten har de svenske forskerne brukt en statistisk metode som gjør det mulig å kontrollere for flere alternative forklaringer, slik at man kan trekke konklusjonen om effekt av tiltaket med større grad av sikkerhet.

Spørsmålet er så: Har Danielsson, Fors og Freij (2011) rett i sine konklusjoner? Svaret er dessverre nei. Grunnen til dette er todelt:

- 1) **Var det ART som ble gitt?** For det første har undersøkelsen ikke dokumentert at det var ART som ble gjennomført, og at implementeringen ble gjort etter boken. Det er velkjent at tiltak som ikke implementeres korrekt kan virke mot sin hensikt eller ha null effekt. Uten en grundigere dokumentasjon av at det faktisk var ART som ble gitt i undersøkelsen, kan man ikke trekke noen konklusjon.
- 2) **Tillater denne undersøkelsen slutning om effekt?** For det andre forutsetter en sammenligning mellom to grupper med tanke på måling av effekt, i dette tilfellet ART vs. kontrollgruppe, at alle andre forhold som kan ha betydning for effekt er kontrollert. Ideelt sett skjer dette gjennom randomisering av deltakerne til betingelsene i undersøkelsen. I dette tilfellet, hvor forskerne retrospektivt har

undersøkt hvordan det gikk med innsatte som gjennomgikk ART vs. ikke gjennomgikk ART, har slik randomisering naturlig nok ikke vært mulig. I slike tilfeller må man være ytterst forsiktig mht hvilke konklusjoner som trekkes om gruppeforskjeller og hvorfor de oppstår. Forskerne har ikke gjort dette i tilstrekkelig grad, og har publisert et negativt resultat det ikke er grunnlag for.

I det følgende utdypes disse to punktene. Begge punktene er helt fundamentale og av avgjørende betydning for kvaliteten på denne undersøkelsen. Hvis rapporten hadde blitt innsendt for vurdering i et vitenskapelig tidsskrift, ville den uten tvil måttet besvare disse to spørsmålene på en tilfredsstillende måte.

1. Var det ART som ble gitt?

Rapporten sier intet om innholdet i de ART-timene som ble gitt. Den eneste informasjonen om implementering som gis, er antall timer intervensjonen har pågått for hhv de som fullførte ART-treningen og de som avbrøt den underveis:

”I medeltal hade programledarna redovisat att de klienter som bedömdes ha fullföljt ART deltagit i 24 gruppmöten om sammanlagt 59 timmar och tre individuella möten. De klienter som var registrerade som att de hade avbrutit programmet hade i genomsnitt genomgått åtta gruppmöten om sammanlagt 18 timmar och två individuella möten. ⁴¹ Majoriteten av klienterna (402 personer eller 77%) hade genomgått ART på anstalt.”

Samlet 523 personer (menn) gjennomgikk ART, av disse fullførte 309 mens 214 avbrøt behandlingen.

Bengt Daleflod har i en mailutveksling med Ingrid Freij spurt hvordan ART-intervensjonene ble implementert, og Freij svarer (mail 14. februar 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 samliga klienter som bedömdes ha fullföljt.

Dette tydeliggjør at Danielsson, Fors og Freij (2011) har brukt data fra deltakere i intervensjonsprogrammet uten å vite om disse faktisk har mottatt ART, og evt. av hvilken kvalitet. Det store frafallet (214 av 523 deltakere) er i seg selv en god indikasjon på at kvaliteten på intervensjonen har vært dårlig. Ettersom det er velkjent at implementeringen av et tiltak er helt avgjørende for kvaliteten på resultatet (se Barnoski & Aos, 2004), er denne omtrentligheten i omtalen av intervensjonen i rapporten

uakseptabel. Dette poenget er så avgjørende at det alene diskvalifiserer konklusjoner fra rapporten, og gjør det ytterst problematisk å hevde at den sier noe om ART mer generelt.

Så kan man innvende at i en retrospektiv undersøkelse er det ikke mulig å få informasjon av denne typen, men undersøkelsen kan likevel ha verdi. Ja, det kan den, men verdi for hva? Det problematiske med denne undersøkelsen er at den pretenderer å teste effekt av ART brukt i fengselspopulasjonen. Hadde den hatt som siktemål å vurdere de intervensjoner som ble gitt i svenske fengsler i 2003-2006, ville den kunne hatt potensiell verdi som intern evaluering med gyldighet begrenset til dette formålet. Men selv det avhenger av om denne undersøkelsen tillater en slutning om effekt, noe neste punkt vil vurdere.

2. Tillater denne undersøkelsen slutning om effekt?

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, Fors og Freij (2011) har valgt sistnevnte mulighet. En slik gruppesammenligning etter intervensjon er mye brukt, eks. i "gullstandarden" i utprøving av medikamenter og intervensjoner, nemlig randomiserte gruppestudier (RCT).

Problemet i den aktuelle rapporten er at man ikke kunne randomisere (tilfeldig fordele) deltakerne i ART vs. behandlingsgrupper. Dette skyldes at tiltaket allerede er gjennomført (for mange år siden). Deltakerne er dermed inkludert i ART vs. kontrollgrupper på en ikke-tilfeldig måte. Danielsson, Fors og Freij (2011) omtaler (s. 6) studien som en "en kvasiexperimentell design". Dette betyr at undersøkelsen gjennomfører en manipulasjon (ART vs. ikke-ART), men at den ikke har den grad av kontroll som en "ekte" eksperimentell studie har.

I seg selv er dette uproblematisk; det er gjennomført mye god forskning ved bruk av kvasiexperimentelle design. Men samtidig er det velkjent at konklusjoner 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 uvedkommende (counfounding) variabler gjennom randomisering, må disse kontrolleres på andre måter for at en slutning om effekt skal være mulig.

Danielsson, Fors og Freij (2011) har kontrollert for noen utvalgte variabler i utvelgelsen av kontrollgruppen, og de har også kunnet kontrollere for effekt av et par variabler i regresjonsanalysen de har gjennomført. ART-gruppen bestod av alle innsatte som hadde gjennomgått ART i årene 2003-2006, samlet 523 personer. Kontrollgruppen bestod av sammenlignbare personer innsatt i samme periode, samlet 2615 personer. Rapporten

oppgir en del inklusjons- og eksklusjonskriterier for kontrollgruppedeltakerne (som ble hentet fra et stort utvalg av potensielle deltakere, ca. 26000; se rapporten s. 16). Det viktigste kriterium for deltakelse i kontrollgruppen, det at personene i perioden 2003-2006 ikke hadde blitt utvalgt til å gjennomgå ART, berøres ikke i rapporten. Dette faktum gjør imidlertid at gruppene i utgangspunktet er selekterte på en biased (skjev) måte, og gjør det naturlig å undersøke hvilke seleksjonskriterier 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. Hvor mange fikk ellers tilbudet, men takket nei? Hvem ble ikke spurt? I og med at informasjon av denne typen ikke er tilgjengelig, vet vi ingenting om de seleksjons- og selvseleksjonsmekanismer som lå til grunn for inklusjon i ART-gruppen.

Seleksjon til kontrollgruppen ble gjort ut fra krav om at deltakerne skulle være menn (siden alle i ART-gruppen var menn), og at 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 narkotikamissbruk (uppgift saknas/ingen missbruksproblematik/inget gravt missbruk/gravt missbruk) för att matcha behandlingsgruppens andel personer med dessa karaktäristika inom varje grupp.”

Kontrollgruppen ble altså utvalgt for å matche ART-gruppen på to variabler, vold (vold/ikke-vold) og narkotikamisbruk (4 kategorier). I tillegg hadde forskerne tilgang på informasjon om alder, lengde på dom og type domfellelse. Med andre ord: 5 mulige confounding variabler i forhold til effekt av ART-intervensjonen er målt i denne studien. Ingen informasjon om deltakerne utover dette er rapportert eller kjent. Dette betyr at seleksjon av deltakerne, samt analyser av resultatene, er basert på noen ytterst få variabler.

Av de variablene som er målt, er det to – alder og tid i institusjonen – som kanskje er mest interessante i sammenligning av gruppene.

Rapporten (s. 6) dokumenterer at de to gruppene var svært forskjellig aldersmessig:

Samtlige klienter var män och medelåldern var 25,4 år (behandlingsgrupp) och 34,6 år (jämförelsegrupp).

Denne aldersforskjellen er interessant. Den kan indikere at alder og/eller variabler som korrelerer med alder i sin tid har vært et seleksjonskriterium for deltakelse i ART-gruppene. Uansett er det ytterst påfallende at ART-gruppen skiller seg så dramatisk i alder fra kontrollgruppen ved så store utvalg.

Også på den andre variabelen – fengselstid – er gruppene noe ulike, med lengre fengselstid for ART-gruppen:

ART: ”Klienter som hade avtjänat tid på anstalt inom Kriminalvården hade i genomsnitt vistats på anstalt i 1,3 år⁴⁰ (standardavvikelse: 1,0 år; intervall: 0,1-5,3 år).”

Kontroll: ”För de klienter som avtjänat tid på anstalt var den genomsnittliga anstaltstiden 0,9 år (standardavvikelse: 1,1 år; intervall: 0,2-14,0 år).”

Når de to viktigste variabler som er målt i undersøkelsen skiller gruppene, den ene dramatisk, kan man anta at gruppene også på andre relevante variabler vil kunne skille seg fra hverandre. Danielsson, Fors og Freij (2011) har imidlertid ikke rapportert noen slike variabler som kan ha betydning for å belyse forskjellen i tilbakefallsrisiko mellom gruppene.

Det er viktig å merke seg at forskjellene i alder og fengselstid angår målte forskjeller mellom gruppene, uavhengig av intervensjonen. Det sentrale spørsmål i undersøkelsen er likevel om intervensjonen, ART vs. kontroll, faktisk er årsak til en målbar forskjell i tilbakefall etter tiltaket. Det sier seg selv at en konklusjon om dette er ytterst vanskelig når forskere selv dokumenterer at gruppene er forskjellige på de to viktigste av de målte variablene – til tross for at de har lagt ned mye arbeid i å få gruppene så like som mulig.

I vurderingen har jeg ikke gått inn på de statistiske analysene som rapporteres. Resultatene rapportert i Bilag 1 (s. 31) indikerer nokså klart at ART-gruppen kommer dårlig ut. Spørsmålet er om dette bildet kan endres når mulig confounding variabler trekkes inn i analysen. Forfatterne diskuterer problematikken rundt confounding variabler slik (s. 18):

”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 återfallsrisk. 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.

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).

Merk at analysen her beskrives som en metode som ”...tar hänsyn till andra påverkansfaktorer...”. Problemet er bare at forskerne har hatt et svært begrenset antall variabler å legge inn i sine analyser. For å kunne legge en mulig confounding variabel inn

i analysen, må den faktisk være målt. Alder og tid i institusjon, som forskerne har lagt inn i analysene, er interessante nok, men man kan ikke på det grunnlag si at man i analysene har "tatt hensyn til andre påvirkningsfaktorer" – det har man simpelt hen ikke gjort. Hvis denne undersøkelsen hadde vært utført som en ordinær kvasi-eksperimentell undersøkelse som kontrollerer 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. Her kan vi liste opp en rekke interessante variabler relatert til personlighet, motivasjon, utdanningsnivå, familierelasjoner. Ingen slike foreligger i denne undersøkelsen.

Til slutt litt om design og slutning om kausalitet (indre validitet): Et kvasi-eksperimentelt design bør 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, og en slik ville heller ikke vært mulig gitt den måten den avhengige variabelen måles på her (en dikotom registrering etter løslatelse, lovbrudd ja/nei). 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 det ikke ut til å ha foregått noe detektivarbeid i det hele tatt.

Konklusjon

Begge de diskuterte momentene peker mot den konklusjon at denne undersøkelsen ikke er holdbar. Vi vet ikke hva forskerne måler effekt av, og vi vet ikke om den påviste forskjellen mellom gruppene – økt risiko for tilbakefall i ART-gruppen – skyldes intervensjonen eller noe annet. Disse metodiske poengene ville en reviewer rutinemessig ha påpekt om forskerne hadde forsøkt å få denne rapporten publisert i et vitenskapelig tidsskrift. Denne undersøkelsen har verken et negativt resultat eller et nullresultat; den har et ikke-konklusivt resultat.