With financial support from the DAPHNE Programme of the European Union.

Mulighed for samarbejde mellem de Europæiske lande, i forbindelse med deres evalueringsstudier inden for programmer som arbejder med personer som anvender sig af vold i hjemmet

Arbejdsdokument nr. 3 fra Daphne III projektet "IMPACT: Evaluation of European Perpetrator Programmes (Programmet for evaluering af Europæiske udøvere af krænkende adfærd)"

Christian Scambor, Katarzyna Wojnicka & Elli Scambor

Liste over ofte benyttede forkortelser

CBT … Kognitiv adfærdsterapi
CCR … Koordineret svar fra samfundet
DVPP … Program for udøvere af vold i hjemmet
RCT … Tilfældige kontrollerede forsøg
T0 … Tidspunkt for observation: Ved henvisning, start af programmet
T1 … Tidspunkt for observation: "Før", ved starten af programmet (f.eks. grupperådgivning)
T2 … Tidspunkt for observation: "Under", i løbet af programmet (f.eks. grupperådgivning)
T3 … Tidspunkt for observation: "Efter", ved slutningen af programmet (f.eks. grupperådgivning)
T4 … Tidspunkt for observation: "Opfølgning", dette sker en vis periode efter programmet er afsluttet (f.eks. 6 måneder efter grupperådgivningen er afsluttet)
IMPACT ... Daphne III projekt "Evaluation of European Perpetrator Programmes (Programmet for evaluering af europæiske udøvere af krænkende adfærd)"
WAVE … Women Against Violence in Europe (Kvinder mod vold i Europa) (international organisation)
WWP … Arbejde med udøvere af vold
WWP – EN … Work With Perpetrators – European Network (Europæisk Netværk for Arbejde med udøvere af vold) (international organisation)

This publication has been produced with the financial support of the DAPHNE Programme of the European Union. The contents of this website are the sole responsibility of Dissens - Institute for Education and Research e.V. and the project partners and can in no way be taken to reflect the views of the European Commission.

1. Introduktion

I henhold til eksperterne inden for dette område, er evaluering af programmer som arbejder med udøvere, meget vigtigt for de praktiserende behandlere, samt programmets ansatte, for at de kan overvåge og forbedre kvaliteten af deres arbejde; samt hjælpe politikerne med at udvikle strategiske beslutninger (Geldschläger et al., 2014; Hester, Lilley, O'Prey, Budde, 2014). Desværre er der op til i dag, kun offentliggjort nogle få Europæiske programmer i videnskabelige magasiner – selvom 2, Hester et al. (2014) i deres gennemgang af Arbejdsgang 2, har fundet frem til minimum 65 nationale studier (offentliggjort i magasiner eller som "gråzone litteratur"). Disse studier virker dog fragmentariske. Størstedelen af disse studier blev udført på et enkelt sted, og der var blot nogle få af dem som kan beskrives som værende flersidede evalueringsprojekter. Generelt vil rapporterne bestå af nationale studier med forskellig tilgang, forskellige teorier når det kommer til ændringer, metoder, værktøjer osv. Resultater fra Arbejdsgang 2 giver indtrykket af, at evalueringen af de Europæiske programmer for udøvere er både rodet og komplekst, og det er svært at gennemføre analytiske konklusioner af arbejdet med udøverne. Dette resulterer i, at der ikke tages hensyn til den Europæiske specificitet, når det kommer til internationale debatter om forskning og evaluering af programmer med udøvere; dette betyder også "… at eksisterende forskning (f.eks. fra studier fra Nordamerika), ikke uden videre kan sammenlignes med et Europæisk kontekst" (Hester et al., 2014, p. 3).
Derfor er hovedmålene med projektet "IMPACT: Evaluation of European Perpetrator Programmes (Programmet for evaluering af Europæiske udøvere af vold)", som er finansieret af den Europæiske Kommission (Daphne III Programmet), blevet defineret på følgende måde:

> a) Udfyld den manglende kendskab der eksisterer, når der kommer til evaluering af programmer med gerningsmænd i Europa, og
> b) Tilbyde løsninger som går hen imod at harmonisere en overvågning af resultaterne (som et første trin til at harmonisere metoderne for evaluering)

ved at:

- Levere et overblik og analyse af den nuværende metode til at overvåge resultaterne fra de Europæiske programmer til udøvere af krænkende adfærd (Arbejdsgang 1)
- Levere et overblik og analyse af forskningsstudier som evaluerer programmer med udøvere af krænkende adfærd (Arbejdsgang 2)
- Identificer muligheder og forhindringer for resultater af forskningsstudier fra flere Europæiske lande (Arbejdsgang 3)
- Udvikle en samling værktøjer og bedste praksis retningslinjer til at måle resultaterne (Arbejdsgang 4)

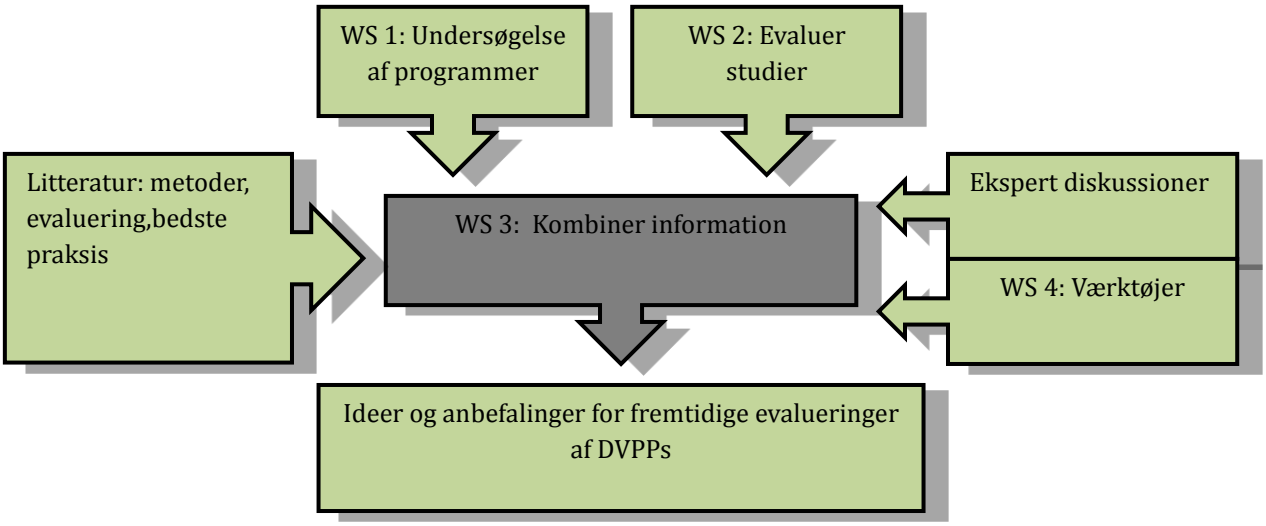
Dette arbejdsdokument præsenterer den metodik samt resultater af de aktiviteter som er foregået i Arbejdsgang 3. Identifikationen og forhindringerne for resultater af forskningsstudier fra flersidet/flere Europæiske lande var baseret på:

- Det (foreløbige og endelige) resultat af Arbejdsgang 1: Overblik af praksissen med at overvåge resultatet (meningsmåling)

- Det (foreløbige og endelige) resultat af Arbejdsgang 2: Overblik af de Europæiske evalueringsstudier
- Gennemlæsning af den relevante litteratur inden for området af forskningsmetodik og evaluering
- Interviews og diskussioner som er foretaget med eksperter og praktiserende behandlere indenfor området med arbejdet med udøvere, samt samarbejdende organisationer og instanser.

På basis af resultaterne fra meningsmålingen i Arbejdsgang 1, er der blevet stillet spørgsmål til, i hvilken udstrækning sammenlignelige metoder og værktøjer til at måle resultater eller evalueringer allerede er i brug i de forskellige organisationer, samt hvilke nationale og institutionelle forudsætninger der kan identificeres til at fremme, eller forhindre, multinationale evalueringsprojekter. Bygget på det kendskab som er opnået ved at gennemlæse forskningen i Arbejdsgang 2, er der indsamlet metodiske elementer som viser en stor sandsynlighed for at de kan benyttes i forskellige nationale sammenhænge. På basis af dette, vil Arbejdsgang 3, arbejde med muligheden for evalueringsprojekter som inkluderer mere end en location, og mere end et land. Aktiviteterne i Arbejdsgang 3 er blevet forbundet med Arbejdsgang 4, for at informere om udviklingen af "værktøjssamlingen", så disse kan bruges til at harmonisere den praksis som er relateret til målingen af resultaterne i programmerne, og dermed hjælpe med til det grundlæggende arbejde af evalueringsprojekter fra flere lande.

Figur 1. Position og funktion af Arbejdsgang 3 inden for IMPACT-projektet.



Bemærk. WS… Arbejdsgang. DVPPs… Program for udøvere af vold i hjemmet.

I dette arbejdspapir er arbejdet som er foretaget i Arbejdsgang 3 blevet opsummeret med speciel fokus på bidrag fra de mange eksperter, samt vores konklusion for fremtidige evalueringsprojekter inden for området hvor der arbejdes med udøvere. En mere detaljeret analyse af de problemstillinger som diskuteres i dette arbejdsdokument vil blive udleveret i to separate videnskabelige artikler.

Forfatterne vil gerne takke alle partnere og tilknyttede partnere til IMPACT projektet, i særdeleshed Heinrich Geldschläger og Marianne Hester samt deres hold, som har været med til at levere de vigtige og foreløbige resultater som har hjulpet med at komme videre med Arbejdsgang 3. Vi vil også gerne takke alle de eksperter som har deltaget i interviews og diskussioner.

2. Metoder

Hovedmålet med Arbejdsgang 3, var at identificere muligheder og forhindringer ved multi-lokaliseret/flere Europæiske landes evalueringsprojekter angående programmer som benyttes i arbejdet med mænd som udviser en voldelige adfærd i hjemmet, overfor kvinder og børn i deres umiddelbare nærhed (DVPPs – Programmer for personer som udøver vold i hjemmet). Den tilgang vi benyttede, til at identificere evalueringsmodeller med mulighed for at kunne implementeres i forskellige Europæiske lande, samt dets kontekst bestod af:

(1) analyse af den eksisterende litteratur i forbindelse med evaluering af programmer med udøvere af krænkende adfærd (både i forhold til Europa og andre lande), samt anden relevant litteratur,
(2) ved at benytte resultaterne fra Arbejdsgang 1 og Arbejdsgang 2, kombineret med yderligere analyser af de data som var indsamlet i disse to Arbejdsgange, og
(3) udføre interviews og diskussioner med internationale eksperter og praktiserende behandlere, indenfor området af behandling af udøvere som udviser en voldelig adfærd i hjemmet.

Der skal lægges vægt på, at fokus i Arbejdsgang 3 var et (a) *resultat* (ikke en *proces* eller andre aspekter i forbindelse med vurderingen af arbejdet på udøverprogrammer), og (b) en *evaluering* med følgende fokus på *forskningsprojekter,* som går længere end den måling som foretages i den daglige brug af disse programmer. (Ikke desto mindre er disse aspekter forbundet, og bør oftest tages i betragtning sammen: Et resultat afhænger af de underliggende processer; og ideelt, hvis den praksis der benyttes til at evaluere resultatet er kompatibel med, eller endda en del af den metodologi som bruges i disse evalueringsprojekter.)

2.1 Analyse af litteratur

De mange resultater som er fundet i offentliggjorte artikler, har gjort det muligt at forberede et internt overblik over de mest almindelige evalueringsmodeller, værktøjer, strategier og metoder, og de har været med til at udforme de spørgsmål som er brugt i interviews med vores ekspert fokusgruppe, samt ekspert diskussioner. Nogle af de kilder som blev benyttet til at finde frem til denne vigtige information, var bl.a.:
Batterer Intervention Systems: Issues, Outcomes, and Recommendations (Gondolf, 2002); *Domestic Violence Perpetrator Programs in Europe , Part I: A Survey of Current Practice; Part II: A Systematic Review of the State of Evidence* (Hamilton, Koehler & Lösel, 2013; Akoensi, Koehler, Lösel & Humphreys, 2013); *Changing Violent Men* (Dobash, Dobash, Cavanagh & Lewis, 2000); *Violent Men and Violent Context* (Dobash & Dobash, 1998); *Evaluating Criminal Justice Interventions for Domestic Violence* (Dobash & Dobash, 2000); *Rethinking Survey*

Research on Violence Against Women (Johnson, 1998); *Description of a Voluntary Attendance Program for Abusive Men* (Payarola, 2011); *Interviewing Violent Men: Challenge or Compromise?* (Cavanagh & Lewis, 1996); *Men Researching Violent Men: Epistemologies, Ethics and Emotions in Qualitative Research* (Cowburn, 2013); *Encountering Violent Men: Strange and Familiar* (Gottzen, 2013); *Batterer Intervention Programs: A Report From the Field* (Price & Rosenbaum, 2009); *Counting Men: Qualitative Approaches to the Study of Men and Masculinities* (Patulny & Pini, 2013); *World Report on Violence and Health* (WHO, 2002); samt to arbejdsdokumenter fra IMPACT projektet: *Outcome Measurement in European Perpetrator Programmes: A Survey* (Geldschläger, Ginés, Nax & Ponce, 2014); *Overview and Analysis of Research Studies Evaluating European Perpetrator Programmes* (Hester, Lilley, O'Prey & Budde, 2014).

En anden vigtig informationskilde i forbindelse med Arbejdsgang 3, var den metodiske litteratur omhandlende evaluering (f.eks. Bamberger, Rugh & Mabry, 2006; Rossi, Lipsey & Freeman, 2004; Shaddish, Cook & Campbell, 2002; Stufflebeam & Shinkfield, 2007). Der blev udarbejdet en specifik liste med udtryk og ord fra denne evalueringslitteratur, som blev udleveret til projektgruppen for at sikre, at der blev benyttet den samme form for sprogbrug gennem hele projektet.

Sideløbende med IMPACT projektet, er der blevet gennemført et betydningsfuldt evalueringsprojekt i Storbritannien, som hedder MIRABAL1. Dette projekt har som mål, at besvare nogle af de spørgsmål som også er vigtige for IMPACT projektet. Projektpartneren Respect (UK) har en førende rolle i MIRABAL projektet, og der blev derfor udvekslet vigtige informationer mellem disse to projekter. Kontakter og diskussioner mellem disse to projekter har været en værdifuld informationskilde (f.eks. ved workshoppen i Bristol i Storbritannien i 2013, hvor Nicole Westmarland præsenterede de nyeste udviklinger og metodologier fra MIRABAL projektet). De mange artikler som til dags dato har været offentliggjort i forbindelse med MIRABAL projektet, har været en værdifuld informationskilde (Alderson, Kelly & Westmarland, 2013; Alderson, Westmarland & Kelly, 2012; Downes, Kelly & Westmarland, 2014; Phillips, Kelly & Westmarland, 2013; Westmarland & Kelly, 2012; Westmarland, Kelly & Chalder-Mills, 2010). Resultaterne fra MIRABAL projektet vil blive præsenteret i begyndelsen af 2015, og dette projekt vil med garanti give nye impulser til fremtidige evalueringsprojekter som vil tage sted på flere lokalisationer. Det skal også bemærkes, at den Amerikanske ekspert Edward Gondolf (2012), har benyttet MIRABAL projektet som et eksempel til en ny udvikling inden for evalueringsforskning som han kalder "praktiserende evaluering" (side 81), som skal løse problemerne med de mangler som opstår, grundet en ukorrekt akademisk tilgang til evaluering af programmer med udøvere.

2.2. Anvend resultaterne fra Arbejdsgang 1

Fokus i Arbejdsgang 3 har været på evalueringsmodeller, f.eks. udvikling, værktøjer og procedurer inden for evalueringer. I Arbejdsgang 1 blev der udført en meningsmåling for at indsamle information om de Europæiske programmer inden for området med det arbejde der blev udført med udøvere af vold i hjemmet, samt praksissen med at måle resultaterne af dette arbejde, og udføre en evaluering af deres arbejde. Som et resultat af dette, havde vi forventet at opnå information om lovende praksisser som allerede var i brug, eller som blev brugt i evalueringsprojekter, som ville vise en mulighed for at disse kunne bruges andre

steder. Selvom forfatterne udtaler at, "… der er ikke fokus på et videnskabeligt forskningsresultat, men på overvågningen af det daglige udfald udøverprogrammer udløser, som en del af deres ydelser", (Geldschläger et al., 2014, p. 4), forventede vi at få visse hentydninger fra Arbejdsgang 1, som vi kunne benytte til planlægning af fremtidige evalueringsstudier. Ved at gennemgå resultaterne fra Arbejdsgang 1, ønskede vi at udforske om der var en fællesnævner for elementer som allerede var i brug, samt tilgange som kunne benyttes til fremtidige studier.

2.2.1. Vigtige forhold fra Arbejdsgang 1

Den udførte meningsmåling i Arbejdsgang 1 indeholder data fra 134 programmer fra 22 lande, og resultatet af denne meningsmåling er blevet offentliggjort i et arbejdsdokument (Geldschläger et al., 2014). De vigtigste resultater og konsekvenser for Arbejdsgang 3 er opsummeret nedenfor. Hvis der ønskes yderligere detaljer, henvises læseren til Geldschläger et al. (2014).

Henvisninger

De fleste programmer som arbejder med udøvere af krænkende adfærd, arbejder med mænd som enten kommer af sig selv, eller som er blevet henvist gennem en anden instans. Visse af disse programmer vil udelukke nogle former for henvisninger, eller fokuserer på en bestemt form for henvisninger (22% af programmerne arbejder kun med mænd som henvises gennem retssystemet, og 19% arbejder slet ikke med mænd som henvises gennem retssystemet; Geldschläger et al., 2014, p. 12). Mange programmer har fokus på mænd som henvises gennem retssystemet eller som kommer af sig selv (eller fra andre instanser) – men størstedelen af disse programmer arbejder med mange forskellige former for henvisninger. Fremtidige evalueringsstudier fra flere lokalisationer skal huske denne forskel.

Samarbejde

Næsten alle programmerne (98%) som var en del af denne meningsmåling, samarbejder med en form for partner-organisation, nogle af de vigtigste er: socialforsorgen, politiet, vidnebeskyttelse, retssystemet, behandlingssteder for alkohol- og stofmisbrug. 75% af programmerne udtaler, at de er en del af et samarbejde mellem institutioner mod vold i hjemmet, og en del af disse programmer tilbyder også støtte til de udsatte indenfor selve organisationen. Derfor kan det konkluderes, at programmerne i denne meningsmåling, følger en vis "systemtilgang", som kan ses i mange lande. Typen og kvaliteten af disse netværk kan være forskellige fra land til land, og burde tages i betragtning i udformningen af et multi-lokaliseret evalueringsprojekt. Disse resultater betyder dog også, at det er muligt at fokusere på de programmer som allerede er en del af et samarbejde med andre instanser. De programmer som har en anden tilgang end denne, er dog også interessante – de kan være en del af et evalueringsprojekt, og herefter studeres i andet trin, eftersom de kræver en anden tilgang og udvikling.

Optagelse

Næsten alle programmer (95%), har en vis form for optagelsesprocedure: fra en enkeltstående kontakt til en længere individuel rådgivning eller bedømmelsesfase i løbet af denne optagelsesfase. Dette er basalt set en nyttig betingelse for at gennemføre evalueringsprojekter: for mange programmer er implementeringen af en vis form for indsamling af data i starten, en "normal" procedure. Omkring to tredjedel af programmerne benytter sig af visse former for værktøjer i optagelsesfasen – men der er ikke nogen dominerende værktøjer. Et bredt udvalg af spørgeskemaer, manualer samt andre værktøjer nævnes. Her er der behov for en vis form for harmonisering – men det er også vigtigt, at de fleste projekter som *udgangspunkt* arbejder med værktøjer. Fremtidige evalueringsprojekter kan opbygges på disse kompetencer.
Størstedelen af programmerne kan tilføje en form fort inklusion / udelukkelses kriterium, som er baseret på optagelsesproceduren (f.eks. nævner en tredjedel "alkohol-/stofmisbrug" og "mentale lidelser" som udelukkelseskriterier). Dette betyder, at størstedelen af programmerne kan foretage udtalelser om deres klienters karaktertræk, i forhold til visse opførsels- og psykologiske karaktertræk. Dette er endnu en vigtig ting at huske ved evalueringsstudier som foregår ved flere locationer: det er muligt at implementere procedurer for indsamling af data før programmet starter, for at få mulighed, senere hen, for at studere forskellighederne mellem forskellige typer af klienter, samt dem der dropper ud af programmet.

Rammer

Omkring to tredjedele af programmerne benytter sig af grupperådgivning (kombineret med andre former for arbejde/rådgivning); der er også ofte mulighed for individuel rådgivning, men grupperådgivning er den mest fremherskende arbejdsmetode. Når det kommer til multilokaliserede evalueringsstudier, vil begge fremgangsmåder (grupperådgivning, individuel rådgivning), være interessante. Andre fremgangsmåder ses mere sjælden (f.eks. par rådgivning, meditation), disse fremgangsmåder kunne man eventuelt inkludere i vise evalueringsformer, men de vil dog kræve specifikke evalueringsmodeller.
Når det kommer til behandlingens "dosering", er der mange variationer inden for de forskellige programmer, med mellem 14 og 52 sessioner for størstedelen af programmerne, som svarer til en varighed på 14 til 52 uger. Denne variation er vigtig for disse evalueringsstudier når det kommer til varigheden, hvilket betyder, at man f.eks. kan sammenligne korte og længerevarende programmer.

Arbejdstilgang

Når det kommer til arbejdstilgangen i mange af de programmer som arbejder med udøvere af krænkende adfærd, vil de fleste programmer bevæge sig inden for området af kognitiv adfærdsterapi (CBT), samt en psykoedukativ tilgang. Hvis man kombinerer CBT, psykoedukativ eller en Duluth lignende tilgang, vil to tredjedel til tre fjerdedele af programmerne benytte sig af disse fremgangsmåder. Disse fremgangsmåder er selvfølgelig ikke identiske – men de har visse træk til fælles, især når det kommer til teorien (teori om indlæring). For USA, påpeger Gondolf (2012) to store tilgangskategorier, i arbejdet med udøvere: c-behavioristisk og psykodynamisk tilgang (også kendt som den "nye psykologi" tilgang), hver tilgang er i sig selv meget forskellig. Når det kommer til de Europæiske studier, kan vi også se en variation af tilgange, hvor der nogle gange kombineres elementer fra forskellige tilgange.

Måling af resultater

Når det kommer til at måle resultaterne, udtaler 81% af programmerne, at de måler deres resultater i en hvis grad (side 31). De vigtigste kilder er klienterne selv (94% af dem som sagde at de målte deres resultater), facilitatorer (63%), samt partnere eller tidligere partnere (60%), men også andre servicecentrer (54%), rapporter fra politiet/domstolene (43%), samt støttecentre for udsatte (32%), var inkluderet. Omkring to tredjedele af programmerne meldte, at de også benyttede sig af opfølgninger.
Dette resultat viser, at det at indarbejde andre informationskilder end de mænd som deltager i programmet, er en praksis som benyttes ofte og som er udbredt blandt disse projekter; dette er noget som fremtidige evalueringsstudier kan bygge videre på. *Partnernes perspektiv* ses som den vigtigste informationskilde i *evalueringsprojekter,* og i de fleste lande, kan der findes programmer som inkluderer information fra partnere/tidligere partnere. Ifølge informationerne fra de programmer som deltog i meningsmålingen findes der lande, som f.eks. Storbritannien, hvor dette er udbredt, og andre hvor det ikke er så udbredt (f.eks. Frankrig, Schweiz). Der bør også tages hensyn til *programmer inden for retssystemet* i visse lande, især de "… 'domstols bemyndiget' programmer der i praksis ikke benytter sig af information fra partnere eller støttecentre for udsatte, som forventet, der en del af disse programmer indikerer at de ikke kontakter partnerne… og måler ikke resultatet med partnerne …" (Geldschläger et al. 2014, s. 46).
Vi har dog fundet programmer i 17 lande fordelt på hele kontinentet, som benytter sig af informationer fra partnere/tidligere partnere, når arbejdets resultat måles. Vi kan derfor konkludere, at der ikke burde være noget grundlæggende problem med at udføre studier i flere lande, som inkluderer partnerne, når det kommer til at måle resultatet af programmet. De steder hvor dette ikke er muligt, vil de respektive programmer få en anden rolle som en del af fremtidige evalueringsstrategier.
I forhold til de *mændene i programmerne,* var de hyppigste målt resultater følgende: Ikke voldelighed/ reducering af vold, holdninger og meninger, klientens kommunikationsevner, samt en reducering af risikoen for voldelige adfærd. Psykologiske aspekter, faderskabsevne og livskvalitet blev også målt af halvdelen til to tredjedele af programmerne.
De vigtigste dimensioner som *partnere/tidligere partnere* blev spurgt til, var: vold, tryghedsfølelse, reducering af risikoen for vold, børns sikkerhed, livskvalitet, mandens attitude, mandens kommunikationsevner, samt mandens faderskabsevner.
Der findes i det hele taget en stor variation af metoder, hvorpå resultatet af udøverprogrammerne kan måles. "En af de vigtigste konklusioner i denne meningsmåling er derfor, at de Europæiske programmer for udøvere af vold, langt fra er enstemmige omkring benyttelsen af standardmetoder for at måle resultater, og der er et tydeligt harmoniseringsbehov. " (Geldschläger et al., 2014, p. 56) I Arbejdsgang 4 af dette IMPACT projekt, vil vi bevæge os i den retning.

2.2.2. Konklusion

Vi konkluderer ud fra resultaterne af meningsmålingen i Arbejdsgang 1, at der ikke foreligger væsentlige problemer i at måle resultaterne i flere lande og flere programmer på en forståelig måde, dette inkluderer også hvordan mændene har opfattet programmerne, partnere og børn, officielle rapporter, facilitatorer – der kræves dog ekstra ressourcer i mange af programmerne, for at være en del af et *evalueringsprojekt. 2* behandlere synes

åbne for disse evalueringer, og diverse dataindsamlings procedurer benytte. Ikke desto mindre, breder evalueringsstudier sig til langt mere end at måle og overvåge resultater. Derfor bør multilokaliserede evalueringsstudier udøves som separate projekter med ekstra finansiering. På baggrund af resultaterne fra Arbejdsgang 1, finder vi det ikke muligt, at gennemføre evalueringsstudier på flere lokalisationer / flere lande, baseret på programmernes nuværende praksis, dette skal ses i lyset af den nuværende praksis i mange af disse programmer, i forbindelse med måling af resultater eller indsamling af informationer i optagelsesfasen. De eksisterende praksisser er ganske enkelt for forskelligartede til, at de kan benyttes som et studie med flere lokalisationer. Ekstra procedurer til disse evalueringsstudier vil kræve ekstra ressourcer for at sikre at alle programmer kan være en del af studiet.

2.3. Anvend resultaterne fra Arbejdsgang 2

I Arbejdsgang 2 blev fokus ændret fra praksis til forskning. Formålet var her at opnå kendskab til de fremgangsmåder som bruges i evalueringsforskning inden for programmer med gerningsmænd i Europa. Forskningsholdet identificerede studier som var offentliggjort i videnskabelige magasiner, samt "grå litteratur", på både engelsk og andre europæiske sprog. Metoder, udvikling, måling af resultater samt yderligere sider af disse studier blev analyseret. Dette betyder, at resultaterne fra Arbejdsgang 2 var meget vigtige for målet i Arbejdsgang 3 (som er at give anbefalinger for fremtidige evalueringsstudier angående evalueringsmodeller).

2.3.1. Vigtige forhold fra Arbejdsgang 2

Hyppighed i forhold til land

Hester et al. (2014) beskriver i detaljer deres tilgang til identificering af eksisterende studier. Der blev fundet fire meta-analyser og sammenstillinger, som var baseret på et samlet antal af femten studier. Det er interessant at se, at størstedelen af disse studier kom fra Storbritannien (7), og Spanien (5), disse to lande var også dem der havde flest programmer identificeret i Arbejdsgang 1.
Kombineret med andre trin til at identificere evalueringsstudier, både offentliggjorte og ikke offentliggjorte (eller offentliggjort som grå litteratur), blev der fundet frem til yderligere 65 studier som kunne bruges til en mere gennemgribende analyse i Arbejdsgang 2. Endnu engang kunne man se, at mange af disse studier kom fra Spanien og Storbritannien, samt fra centraleuropæiske lande. Vi kan se en overlapning af hyppigheden når der kommer til *programmer* (Arbejdsgang 1), samt *evalueringsstudier* (Arbejdsgang 2) som giver mening: lande som har flere (synlige) programmer til udøvere, har en tendens til at udføre og offentliggøre flere evalueringsstudier.

Problemer i forhold til studierne

Der var et bredt udvalg af problemer i de analyserede evalueringsstudier: Nogle af de vigtigste er vist nedenfor:

- *Design:* Når det kommer til udviklingen af forskningen kunne man stort set ikke finde nogen Tilfældige Kontrollere Prøver (RCT). Ud af n=58 studier fra de fem Europæiske regioner som havde fokus på *resultatet* (de var fokuseret på resultatet, eller på resultatet og processerne), var der kun 2 som havde en basis for RCT, og disse to var studier som var foretaget i et fængsel og på en psykiatrisk klinik.
- *Kontrolgrupper:* Mange af studierne mangler en kontrolgruppe, eller de benyttede kontrolgrupper er for små. Dette faktum er en meget stor begrænsning af de tilgængelige studier som en kontrolgruppe (eller tilsvarende form for kontrol som tillader tilfældig slutning), som er vigtige for den interne gyldighed af ethvert studie.
- *Terminologi:* Disse studier viste, at der blev brugt et bredt udvalg af definitioner når det kom til vigtigt sprogbrug, (f.eks. ”start”, ”færdiggørelse”). Der eksisterer ikke et fast sprogbrug.
- *Tidspunkter:* Der var forskelle i tilgange og metodologier, når det kom til hvilket tidspunkt de forskellige data blev indsamlet på. Hester et al. (2014) har præsenteret et skema til deres analyse, som skulle bruges og fremmes for en fælles forståelse af observeringstidspunkter i disse evalueringsstudier. Skemaet består af fem forskellige tidspunkter, hvorpå der skal indsamles data til en evaluering (T0: Før start; T1: Start; T2: Under forløbet; T3: Slutningen; T4: Opfølgning). Dette skema er beskrevet nærmere nedenfor.
- *Naturlig afgang:* Raterne for naturlige afgange var ikke tydelige nok, eller det var ikke tydeligt i hvilke faser at denne afgang foregik. I de fleste tilfælde manglede der information om størrelse/antal samt afgang på *hver* fase. Den information som findes varierer meget, både når det kommer til eksempler samt størrelse/antal – dette skyldes en manglende standard og etableret sprogbrug når det kommer til evalueringsstudier i forbindelse med programmer som arbejder med udøvere af krænkende adfærd.
- *Henvisning:* Det er vigtigt at have information om henvisninger, eller måden som programmet modtager dens klienter på (kommer de selv eller gennem retssystemet, anden måde af henvisning). Det er også vigtigt, at der er information vedrørende hvorfor klienten forlader programmet, demografisk information, samt enhver anden tilgængelig information (klienter som henvises inden programmet starter, frafald, klienter der fuldfører programmet). I mange tilfælde kunne man se at denne information manglede.

Måling af resultater og indvirkningen

Når det kommer til resultaterne af disse målinger, kan man se, at alle regioner benytter sig af data fra udøverne, officielle data fra retssystemet, samt data fra de udsatte, undtagen de sydlige regioner, hvor der ikke forefindes data fra de udsatte i de tilgængelige studier. Der er dog visse programmer fra de sydlige regioner som indeholder perspektivet fra de udsatte, som set i Arbejdsgang 1. Det ser dog ud til, at der er en vis forskel på praksisser og de juridiske muligheder, når der kommer til programmer som benyttes i fængsler, i forhold til programmer som benyttes i det almene samfund. Programmer som udføres indenfor retssystemet tager ikke partnernes perspektiv med, men de er dog evalueret oftere end andre lokale programmer. De benytter sig også af en del validerede/standardiserede psykometriske værktøjer som fokuserer på de psykologiske aspekter og ændringer med mandlige udøvere. Dette resulterer i en stærk psykologisk tilgang til evalueringen af programmer med udøvere inden for retssystemet i de sydlige regioner, hvor de fleste studier kommer fra Spanien.

De mest omfattende resultater af disse målinger er fundet i de vestlige regioner, "… inklusiv ændringer i holdninger mod kvinder samt vold mod kvinder, officielle og selv-rapporterede recidivisme, gentagende forulempelse, livskvalitet, sikkerhed og velfærd for disse kvinder/udsatte og deres børn, niveau af stress over forældrerollen, samt gerningsmandens brug af alkohol/stoffer." (Hester et al., 2014, p.26) Der er fundet frem til, at værktøjer til risikovurderinger benyttes i studier i de vestlige og sydlige regioner.
Generelt kan man sige, at forholdene som studeres, samt de værktøjer der benyttes, varierer meget fra region til region (og fra studie til studie). Denne mangel på overensstemmende metodologi gør det meget svært, om ikke umuligt, at sammenligne disse studier.

Dette betyder, at når der skal gives en overordnet udtalelse om effektiviteten af disse programmer med udøvere, baseret på de tilgængelige evalueringsstudier, kan der kun gives et meget generelt billede: enten vises der statistiske betydelige resultater som peger i retning af positive ændringer; eller der er lovende resultater som ikke har nogen statistisk betydning. Dette mønster kunne ses for mange af måleområderne (ændringer inden for opførsel, holdning, psykologiske, risikofaktorer, motivation og overholdelse). Der er dog begrænsninger når det kommer til resultater inden for indvirkning, i mange tilfælde var der kun mindre testgrupper, høje niveauer af afgang (eller manglende information om afgang), samt en overdriven tillid til mandens udtalelser i deres selv-rapporter i de sydlige regioner.

2.3.2. Konklusion

I forhold til Arbejdsgang 3, kan man sige, at på trods af alle de problemer der blev fundet i de mange studier, så peger resultaterne på, at det er muligt at indsamle og kombinere forskellige data i alle regioner. Der er forskellige fremgangsmåder når det kommer til evaluering, men det virker muligt at tilføje forskellige værktøjer, samt integrere forskellige former for information på tværs af landene (med visse undtagelser, som f.eks. programmer fra fængsler, som spiller en anden rolle i en fremtidig forskningsstrategi for evalueringer).
Hovedproblemet som det ses i dag, virker til at være den uoverensstemmende metodologi som benyttes til evaluering af disse programmer, hvilket gør det svært at sammenligne studierne (udvikling og kontrolgrupper, sprogbrug, værktøjer, data). En harmonisering af denne evalueringsmetodologi er nødvendigt for at kunne overkomme den forskel der findes inden for evalueringsforskning af programmer med udøvere. Dette virker muligt, som forfatterne fra arbejdsdokumenterne i Arbejdsgang 2 også konkluderer:

"Der var nogle interessante forskelle og lighedspunkter mellem og indenfor de forskellige regioner, og det vil være muligt at tage elementer fra de forskellige fremgangsmåder, for på denne måde at påbegynde udviklingen af en solid evalueringsmetodologi." (Hester et al., 2014, s. 39)

2.4. Eksperternes fokusgruppe interviews

Når det kommer til evalueringsstudier af programmer for udøvere af krænkende adfærd, så er der mange interessenter som skal identificeres, som alle har deres egne interesser og perspektiver, disse kan både variere fra hinanden, men de kan også overlappe hinanden:

- *Udsatte som overlever vold i hjemmet:* Den kvindelige partner til en partner som udøver en krænkende adfærd og dennes børn, vil gerne vide om dette program er effektivt nok til at stoppe deres partner/faders vold, samt øge deres sikkerhed.
- *Behandlere indenfor støtte til de udsatte:* De mange professionelle behandlere som arbejder i de støtteorganisationer som hjælper udsatte, vil gerne vide om programmerne er effektive nok til at begrunde en yderligere finansiel støtte til disse programmer, der ellers kunne benyttes til mere støtte til de udsatte.
- *Behandlere i de mange instanser:* Politiet, anklagere, domstolene, kriminalforsorgen samt andre interessenter er også interesseret i at vide om DVPPs har nogle specifikke effekter (i tillæg til eksisterende indblandinger).
- *Behandlere indenfor DVPPs:* Disse professionelle behandlere vil gerne kende resultatet af deres arbejde, og vil have en specifik viden om hvorvidt deres arbejdes er effektivt og volden reduceres, og hvis det er tilfældet, hvordan de kan forbedre deres arbejde for at sikre endnu bedre resultater.
- *Programledere indenfor DVPPs:* De er interesseret i effekten af programmet, samt en forbedring af kvaliteten af disse programmer; dvs. deres interesse ligger i at fremvise effektivitet og forbedre resultater.
- *Offentligheden:* Generelt kan man sige at offentligheden er interesseret i at mindske tilstedeværelsen af volden i hjemmet i vores samfund, samt øge den generelle sikkerhed. Den offentlige diskussion foregår ofte i medierne, som har deres egen agenda. Disse medier vil nogle gange stille de samme spørgsmål som andre interessenter eller grupper, mens andre medier vil arbejde med chokerende og ekstraordinært indhold, som ofte kræver en hårdere straf af de udøvende, samt en strengere lovgivning, der er dog en vis interesse i effekten af programmer som DVPPs.
- *Bidragsydere og politikere:* Interessenter indenfor denne kategori kan ønske vise deres arbejde mod vold til offentligheden, samtidigt med at de sikre at de midler de tildeler programmerne, er effektive og fører ti at opnå forudbestemte mål; samtidig vil de måske ønske at optimere brugen af ressourcerne, f.eks. være interesseret i at besvare spørgsmål vedrørende omkostningseffektivitet.

I afsnittene med *ekspert fokusgruppe interviews* samt *ekspert diskussionerne* (se nedenfor), var spørgsmålet om forskellige interesser samt overlappende interesser af diverse interessentgrupper et problem.

Ekspert fokusgruppe interviews (en social forskningsmetode som kombinerer interviews med eksperter, med interviews med fokusgrupper; se Babbie, 2003; Silverman, 2013), blev udført i oktober måned i 2013, under projektets *Workshop A*, som blev afholdt af Bristol Universitet i Storbritannien. Der blev afholdt fire fokusgruppe interviews som hver varede 45 minutter, med 18 forskellige eksperter (forskere og behandlere som arbejder med gerningsmænd), fra 11 Europæiske lande (Østrig, Tjekkiet, Danmark, Finland, Tyskland, Irland, Polen, Portugal, Norge, Spanien, Storbritannien). I hver gruppe ville facilitatoren (et medlem af projektgruppen), stille tre hovedspørgsmål:

- Hvilke fremmende eller hindrende betingelser ser du for gennemførelsen af et multi-lokaliseret design, på et nationalt eller lokalt niveau?
- Hvilke fremmende eller hindrende betingelser ser du for gennemførelsen af et multi-lokaliseret design, på et inernationalt niveau?
- Hvilke muligheder og forhindringer ser du, i forhold til følgende ide: 'at frembringe et værktøj for behandlere (i udøverprogrammer), at indsamle sammenlignelig data på tværs af lande; ved at sammensætte disse dataer i en pulje, analisere centralt'?

Facilitatorerne forberedte efter disse interviews en intern rapport, som blev brugt som basis for præsentationerne under Workshop B, samt til yderligere analyser (se nedenfor).

Ekspert diskussionen foregik i workshop B som blev afholdt i april måned i 2014 i København i Danmark. Her deltog 41 behandlere, forskere og politikere fra Østrig, Bulgarien, Tjekkiet, Kroatien, Danmark, Estland, Finland, Tyskland, Irland, Italien, Litauen, Holland, Rusland, Spanien, Sverige, Schweiz og Storbritannien; de diskuterede alle problemstillingerne og spørgsmålene som blev udleveret i præsentationen *Mulighederne for evaluering i en fremgangsmåde med flere lokalisationer / lande*:

- Hvilke muligheder og forhindringer kan jeg se i min organisation?
- Hvilke spørgsmål skal besvares i evalueringsstudier:
  - Virker det?
  - I hvilken udstrækning/hvor effektivt?
  - Hvad virker?
  - Specifikke spørgsmål (hvem overfalder igen, hvem har brug for hvad, osv.)

Resultaterne af de diskussioner som foregik i fire grupper blev indsamlet, og blev brugt til yderligere analyser (se nedenfor). Der blev indsamlet yderligere feedback på konceptet af evalueringsstudier på flere lokalisationer / lande i forbindelse med DVPPs, på projektets afslutnings konference som blev afholdt i Barcelona i Spanien (oktober måned 2014).

3. Resultater: Eksperternes perspektiver

Resultatet fra interviews med ekspert fokusgrupperne, samt ekspert diskussionerne, blev opsummeret som en samling af problemstillinger; disse har været en god kilde til information om hvordan man kan arbejde videre med udviklingen af en metodologi for programmer til udøvere. Disse problemstillinger er blevet kategoriseret i fem hovedgrupper: spørgsmål, muligheder, fordele, udfordringer og hindringer.

3.1. Spørgsmål

Et af de vigtigste og mest almindelige stillede spørgsmål, både under interviews og diskussioner, var den grundlæggende forståelse for "programmer til udøvere". I henhold til eksperter fra forskellige lande, så kan definitionen af sådan et program være forskellig: bortset fra de mest almindelige programmer (f.eks. almindelige gruppemøder udført af professionelle facilitatorer, strukturerede, samarbejdende, evaluerende), så blev det følgende nævnt som elementer hvorpå der kunne arbejdes med udøvere af vold i intime parforhold: telefon hjælpelinjer, individuel rådgivning, uformelle møder, frivilligt arbejde og/eller religiøse møder, møder hos AA osv. Derfor bør et af de første trin til at komme videre med et Europæisk evalueringsforskningsprojekt i forbindelse med arbejde med udøvere, være at udarbejde en generel, men samtidig relativ bred ***definition af et program til udøvere***.

Et andet vigtigt spørgsmål var i forbindelse med ***sammenligneligheden af programmer.*** De fleste eksperter har den mening, at det kun er resultater fra sammenlignelige programmer (samme fremgangsmåde, lignende arbejdsmetoder, lignende klienter, længde, størrelse osv.) der bør sammenlignes. Denne indstilling skaber dog tvivl når der kommer til de (mangelfulde) antal Europæiske programmer som var sammenlignelige nok, til at disse kan sammenlignes.

Problemstillingen med sammenlignelighed blev efterfulgt af spørgsmålet om hvilke ***programtyper*** der skulle undersøges. Visse eksperter understregede at fremtidige studier kun burde fokusere på en form for programmer: enten *obligatoriske* programmer eller *frivillige,* mens andre foreslog at begge typer skulle undersøges, eller eventuelt fokusere på ”blandede” programmer (hvor begge former for henvisningstyper eksisterer i programmet).

Endelig talte eksperterne om problemstillingen med ***forskningsproceduren*** af et fremtidigt studie. Konkrete spørgsmål var fremme som f.eks., hvem skulle forskerne være, hvem skulle indsamle data osv. De mange muligheder for og imod blev diskuteret, f.eks. om der skulle benyttes uafhængige eksterne forskere, eller om behandlere som arbejdede med disse udøvere skulle indsamle data, og eventuelt også foretage vise analyser. Når det kommer til motivation og interesse i at evaluere arbejde med udøvere, så kan dette være en fordel senere hen, men der kan også opstå en interessekonflikt som kan vise sig at være en ulempe. I henhold til eksperterne, kan der tales både for og imod begge muligheder (forskellige adgang til data og prøver, udgifter, forskellige evner indenfor analyser, problemet med ”saglighed”, tid osv.). Så beslutningen om hvilken procedure der skal benyttes, skal grundigt overvejes.

3.2. Muligheder

Mange eksperter udtrykte deres interesse i at udføre og støtte op om en form for Europæisk resultatstudie med flere lande / lokalisationer. Som et resultat, blev mulighederne for sådanne studier diskuteret mest. I de relaterede diskussioner, ***programstandarder*** samt ***måling af resultater*** som bruges til evalueringsstudier, var dette særdeles interessant for mange eksperter. Et begyndelsespunkt til at definere minimum standarderne for programmer til udøvere, samt for måling af deres resultater, blev set i *Convention on Preventing and Combating Violence Against Women and Domestic Violence* (“Istanbul Convention”, Council of Europe, 2011), i den korte orientering af kriterier for resultater af Westmarland, Kelly og Chalder-Mills (2010), eller i *Guidelines to Develop Standards for Programmes Working With Male Perpetrators of Domestic Violence* (Work with Perpetrators [WWP] – European Network, 2008). Det skal også bemærkes, at eksperterne er af den mening, at disse standarder skal være baseret på ligestillingsretten.

Ordet ”minimum” blev også nævnt i forbindelse med forsknings- og evalueringsstudier, i en hentydning til at ***minimumsstandarder*** er nødvendigt ***i alle lande-specifikke studier,*** eller hver del af studiet, for at man skal kunne “udtrække” de forskellige afsnit/data. Der blev nævnt flere elementer som allerede er i brug ved forskellige anledninger, f.eks. den miljømæssige model, observeringsmodellen (før, under, opfølgning), bestemte værktøjer (f.eks. oversigter over vold og misbrug). Forskerne er allerede begyndt at beskrive og analysere forskellige relevante problemstillinger, f.eks. Europæisk og national kontekst

(lovgivninger, henvisningsmåder osv.), typer af udøvere, hyppighed og voldstyper, ønskede ændringsområder (udøverens voldelige adfærd, kommunikation, meninger og identitet, familieforhold, osv.). Denne diskussion fik visse eksperter til at stille de første forskningsspørgsmål (uden at give et svar på dette tidspunkt af diskussionen, om hvordan disse spørgsmål skulle håndteres):

- Hvad gør den største forskel når det kommer til at ende vold og øge de udsattes sikkerhed?
- Ændrer mænd sig? Hvorfor? Hvordan?
- Hvad virker for hvem?
- Hvilke (andre) faktorer påvirker en udøvers liv (klasse, religion, baggrund, osv.)?

Til slut talte eksperterne om hvilke forhold der var nødvendige for at forbedre ***kvaliteten af fremtidige evalueringsstudier på flere lokalisationer / lande,*** f.eks. triangulering af forskningsmetoder (kvalitet, mængde, observationer), triangulering af data og teorier, inddrage perspektiver fra de udøvende, de udsatte og behandlere. Endvidere mente eksperterne at det kunne være nyttigt at inddrage samarbejdende organisationer og/eller lokale/specialiserede NGOs (immigrantgrupper, religiøse grupper, osv.).

3.3. Fordele

I alle interviews og diskussioner som blev gennemført, understregede eksperterne fordelene ved evalueringsstudier fra flere lokalisationer. Dette ville være en ***god*** og ***nyttig*** måde, hvorpå man kunne måle volden, samt andre ***resultats målinger*** som skulle bruges eller udvikles til et sådan studie, som ikke kun ville være en fordel for praktiserende behandlere i alle de Europæiske lande, men som også vil være med til at fylde det tomrum som eksisterer når der generelt kommer til kendskab og forskning indenfor vold: i henhold til visse eksperter, er den eksisterende metodologi ikke god nok. Det var især behandlere som udtrykte et håb om, at evalueringen eventuelt kunne støtte ***videreførelsen af de eksisterende programmer,*** idet visse bestemmende instanser forbinder den videre finansiering med de positive resultater fra en evaluering af programmerne.

Endvidere påpegede eksperterne at ideen med at indarbejde perspektivet fra de udsatte, samt organisationer som hjælper de udsatte, ville føre til ***flere omfattende evalueringsstudier,*** inklusiv det ***bredere system*** som var en del af den oprindelige indblanding. Evalueringsstudier som involverer flere Europæiske lande, kan også føre til ***fremningen*** af unikke nationale programmer til udøvere, som endnu ikke er velkendte blandt de forskellige interessenter, endvidere kunne resultaterne fra programmerne virke som en ***videndatabase*** for behandlere som er interesseret i at forbedre kvaliteten af deres arbejde, undgå generelle fejl, osv. Som en ekstra effekt af dette, vil den opmærksomhed som et internationalt studie med arbejde med udøvere vil få, føre til en øget ***samfundsviden*** om programmet, samt hvilken rolle sådanne programmer har i visse lande, og i henhold til visse eksperter, vil dette hjælpe med en mere effektiv fremning på både nationalt og internationalt niveau. Hvis et Europæisk evalueringsstudie med flere lande blev gennemført som et Europæisk projekt, vil dette måske være den eneste mulighed for evaluering af programmer i de lande, hvor der endnu ikke har været evalueret nogen programmer, grundet mangel på ressourcer og metodologi.

3.4. Udfordringer

Eksperterne var enige om at ***ansættelsesproceduren*** var en af de største udfordringer når det kommer til evalueringsstudier i flere lande, især i de tilfælde hvor der ikke er nogen yderligere finansiering. De understregede at et studie som dette, kun ville blive en succes hvis forskerne kunne overtale behandlere, ansatte i programmerne, samt andre vigtige involverede, til at give deres ***tid*** og ***ressourcer*** til at deltage i denne forskning. Deres mening var, at argumentet om at udfylde det manglende kendskab indenfor området med evalueringsstudier af programmer til udøvere, måske ikke ville være nok til at motivere vigtige involverede til at deltage. Derfor er det vigtigt at ligge vægt på andre ***fordele*** som kan komme ud af at deltage i sådanne evalueringsstudier, såsom ekstern træning af programmets ansatte (ikke blot vedrørende evalueringsmetoder), besøg, certificeringer (f.eks. gennem WWP – EN, eller nationale organisationer som Respect i Storbritannien eller BAG Täterarbeit HG i Tyskland). I henhold til de forskere blandt eksperterne, vil en af de mest vigtige fordele ved et evalueringsstudie være feedback fra klienterne, hvilket kan føre til en bedre kvalitet og effektivitet af programmet. I tillæg til dette, kan det at programmet er en del af dette studie, betyde at programmet måske har nemmere ved at opnå finansiering i fremtiden; når først en evalueringsproces er blevet gennemført, og resultaterne viser at programmet har en virkning. Den Europæiske mening om et eksternt evalueringsstudie vil øge troværdigheden af resultaterne på et nationalt niveau.

En anden udfordring er, at finde frem til en ***aftale*** angående bruges af de samme ***evalueringsmodeller*** (kriterium, design, metoder, værktøjer). Eftersom programmerne og de praktiske metoder er forskellige i mange lande, og der på nuværende tidspunkt også bruges forskellige kriterier og værktøjer til at måle resultaterne. Derfor er udfordringen ikke blot at udvælge og benytte de samme værktøjer, men også at vælge de værktøjer som nemmest kan overføres og bruges af alle.

3.5. Hindringer

I henhold til eksperterne, kan udførelsen af et Europæisk forskningsprojekt i flere lande, måske betyde, at der et antal hindringer som skal overvindes. Et af de mest alvorlige problemer vil være ***finansiering*** – eftersom det vil koste en del at udføre et sådan komplekst og omfattende studie. Derfor er det meget vigtigt at der findes finansiel støtte for projektet, men det er også meget udfordrende. Finansiering er ikke kun nødvendigt til at dække udgifterne til et hold af eksterne forskere, samt andre omkostninger i forbindelse med projektet (rejseomkostninger, udstyr, osv.), der skal også være finansiering til at dække det ekstra arbejde som programmernes ansatte udfører, især i de situationer hvor behandlere går til en mindre løn, eller endda arbejder på frivillig basis (hvilket ofte er situationen med programmer som er placeret i tidligere socialistiske lande), og/eller hvor evalueringen af programmet, ikke er en del af den nationale eller regionale finansiering.

Den ***forskellige kontekst*** fra de mange Europæiske lande er en anden hindring, som har indflydelse på både programmets udformning, samt fremtidige forskningsprocedurer. Forskellig juridisk kontekst kan have en indflydelse på valget af specifikke metoder, samt den basale formodning af et forskningsprojekt. De juridiske lovgivninger i forbindelse med opstartsystemet, beskyttelse af personlige data, osv. kan forhindre forskere fra at ***modtage***

bestemte data i en bestemt kontekst. Et af de mest almindelige eksempler på dette blev givet af en ekspert fra Spanien, hvor de ansatte i programmerne til udøvere, inden for retssystemet ikke har adgang til de udsatte, og derfor, kan det være meget svært, og oftest umuligt, at inkludere partnerens perspektiv.

En anden problemstilling er den ***kulturelle kontekst*** af mange af de Europæiske lande, som også kan have en indvirkning på forskningsproceduren; i vise lande er ***forståelsen af vigtige udtryk*** som *vold i hjemmet, de udsattes rettigheder, upassende opførsel, kønsroller,* ligestilling*, osv.* forskellig, og derfor kan det være en udfordring at udvikle forskningsværktøjer som kan bruges i flere lande. Endvidere er der forskel på landenes ***økonomiske*** situation, de forskellige ***religioner*** (i hvilken udstrækning at politikken er påvirket af religion, dette gælder mere eller mindre visse lande), variationen af kulturelle baggrunde i forbindelse med ***indvandring*** osv., men også i forbindelse med ***implementeringen*** af ***ligestilling*** og undgåelse af kønsdiskriminering, graden af implementering af ***programmer til udøvere***, samt niveauet som ses som værende ***acceptabelt for vigtigheden*** af sådan et system.

Når det kommer til programmerne, kan eksperterne se mange forskelle fra land til land, når vi taler om ***teorier for ændringer, kriterier for succes, typer af mænd*** som kan være en del af et program til udøvere, samt forskellige fremgangsmåder når det kommer til ***klientens køn*** i grupperne: der er programmer som kun arbejder med mænd, andre har en mere indlærende tilgang, med forskellige grupper/veje for mandlige og kvindelige udøvere, eller uden at skelne til en forskel mellem disse. Som en opsummering mener eksperterne, at den mest alvorlige (metodologiske) hindring for et Europæisk evalueringsforskningsprojekt i flere lande, faktisk skyldes at ***programmerne til udøverne,*** ganske enkelt er meget ***forskellige***.
Til slut var der nogle eksperter som mente, at det på nuværende tidspunkt ikke vil være muligt at gennemføre et komplekst Europæisk studie, eftersom der ganske enkelt ikke er nok programmer til udøvere (i visse lande som f.eks. Ungarn eksisterer sådanne programmer slet ikke endnu; Kutrovátz, 2013), hvilket kan føre til alvorlige problemer når det kommer til udsnit.

3.6. Hvilke resultater er der kommet fra eksperternes fokusgruppe, samt diskussionen

De mange indlæg fra eksperterne til gruppediskussionerne, samt deres svar til forskellige spørgsmål, er blevet opsummeret til følgende ekspert anbefalinger:

- For det første, skal der udarbejdes en ***definition*** af "programmer til udøvere", som skal benyttes i evalueringsstudier i flere lande. Standardprogrammerne (samarbejde med støttecentre til de udsatte og andre organisationer, baseret på CBT eller en psykoeduaktiv tilgang, med fokus på køn), virker som den bedst mulige tilgang på nuværende tidspunkt; også når det kommer til de metodologiske grunde: vi skal bruge lignende programmer med forskellig kontekst, og ikke forskellige programmer med forskellig kontekst. Programmerne skal også være i stand til at indsamle de nødvendige data for resultater, som afhænger af samarbejde, f.eks. i en planlagt tilgang. Endvidere skal fokus for studiet være på programmer som er ***kompatible*** med eksisterende ***standarder*** og ***politikker*** (f.eks. Istanbul konventionen).
- ***Forskningsprojekt***: Det at arbejde sammen i et nyt finansieret projekt blev set som den bedste måde at realisere et evalueringsprojekt med flere lokalisationer / lande. Der blev stillet spørgsmålstegn til den oprindelige ide med at benyttet værktøjssamlingen til

evaluering, ved at indsamle og analysere data centralt – i det mindste, på nuværende tidspunkt af processen. Ikke desto mindre skal denne værktøjssamling implementeres, og forskning skal opbygges, så vidt som muligt, på denne værktøjssamling; på et senere tidspunkt kan det så være muligt at gå tilbage til den oprindelige ide. Forskellen på denne værktøjssamling (som et resultat af Arbejdsgang 4, til at måle resultater), samt andre værktøjer som en del af en metodologi for evaluering, blev fremhævet af eksperterne. Det at inkludere evalueringsprojekter i det daglige arbejde, blev set som værende meget udfordrende.

- Fremtidige studier skal udvikles på en meget ***omfattende måde,*** inklusiv det ***system*** som er involveret i behandlingen. Den sociale påvirkning af disse programmer kunne udvide det basale design, når det kommer til virkningerne af programmerne. Der skal fra begyndelsen tages hensyn til forskellig ***kontekst*** fra forskellige lande, både når det kommer til juridisk lovgivninger, de mange sider af kontrolsystemerne, samt kulturelle problemstillinger.
- Evalueringskriterier, design, metoder, værktøjer, osv. skal henholde sig til eksisterende ***god praksis indenfor evaluering,*** og tage hensyn til ***eksperternes*** udtalelser.
    - De mange ***spørgsmål*** som skal besvares af disse evalueringsstudier skal være ***systematiske,*** hvis muligt ved at relatere dem til bestemte grupper af ***interessenter***.
- Evalueringsmodeller skal vise en høj ***brugbarhed,*** så de i længden kan indarbejdes i mange projekter i mange forskellige lande.
- Når det kommer til typen af programmer: ***henviste former*** (selv-henviste, eller henvist fra organisationer, blandet) kan bruges som en variabel faktor, ikke som et udvælgelseskriterium.
- Under alle omstændigheder skal der lægges vægt på ***fordelene*** som eksisterer ved at deltage i evalueringsstudier under rekrutteringsprocessen.

4. Sådan går du videre med evalueringen af DVPPs

4.1. Begyndelsespunkt; Her er vi på nuværende tidspunkt

Ved at studere de tilgængelige ressourcer for evalueringer af DVPPS, samt perspektiverne fra eksperterne, er det tydeligt at den oprindelige ide i Arbejdsgang 3 skal tilrettes. Det har ikke vist sig at være så nemt at udarbejde, udvikle og foretage større evalueringsstudier, som kan besvare de mange omfattende spørgsmål som stilles af de mange interessenter, i det mindste ikke på nuværende tidspunkt. I stedet har vi udviklet et koncept for et samarbejde om evalueringer, som skal ses som en række af mindre evalueringsprojekter med lignende metodologier, samt mulighed for samarbejde indenfor disse projekter.

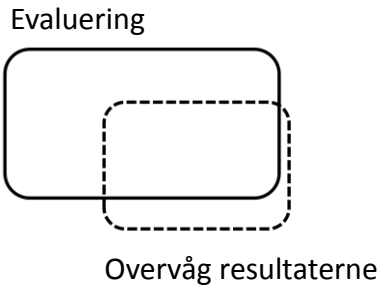
For at opsummere, så er forudsætningen for evalueringsprojekter for flere locationer / lande, det følgende:

Den **primære motivation** for et sådan projekt, er at gøre brug at den megen forskellighed som findes i de nationale reguleringer og ”systemer” (praksis i forskellige instanser, støtte til de udsatte, DVPP, osv.), gennem hele Europa. Denne forskellighed kan være en ressource til evalueringsstudier, en mulighed for at studere DVPPs i forskellig kontekst. Endvidere, eftersom mange af disse programmer endnu ikke har været evalueret, kan evalueringsstudier i flere lokalisationer / lande være en mulighed for at udvide den basale erfaring i forbindelse med DVPP. Som nævnt ovenfor, er der megen mangfoldighed med arbejdet med udøvere fra

land til land, men der kan også findes sammenlignelige programmer (f.eks. CBT, eller Duluth baserede DVPPs), samt visse elementer inden for samarbejdende netværk (CCR). I mange lande er der udviklet og indarbejdet procedurer til at måle resultaterne af arbejdet, men der er megen metodologisk forskellighed, hvilket gør det svært at sammenligne de eksisterende studier. Der er forskelle og problemstillinger i forbindelse med sprogbrug og koncepter (f.eks. ”færdiggørelse”, ”succes”, osv.) samt design (f.eks. i forbindelse med tidspunktet hvor der måles resultater, og de informationskilder der benyttes). Det er svært at oprette kontrolgrupper, og der er etiske bekymringer når man taler om kontrolgrupper; afgang er et stort problem i praksis, dette gælder både for de mænd som deltager, samt for de overlevende. Men, som Hester et al. (2014) konkluderer ud fra deres indsamling af studier, ”… det ville være muligt at tage elementer fra forskellige fremgangsmåder, for at begynde at udvikle en solid evalueringsmetodologi.” (Hester et al., 2014, p. 39)

Eksperterne mener stadig, at det er umagen værd at følge ideen om en tilgang med flere lande, og de forskellige interessantgrupper er interesseret i at få svar på spørgsmål som, programmernes effektivitet inden for de forskellige former for DVPPs. Som forskellige interessenter har udtalt, skal der være fokus på programmer som er kompatible til eksisterende standarder og politikker (f.eks. Istanbul Konventionen). Det er dog tydeligt at evalueringsstudier som er i stand til at besvare de relevante spørgsmål, kræver en masse arbejde og ressourcer, og mange projektledere og professionelle er skeptiske over at skulle indarbejde dette i deres daglige arbejde. Sagt på anden måde, det virkede ikke muligt at begive sig i retning af internationale evalueringsstudier med flere lokalisationer / lande, uden ekstra ressourcer. Mange interessenter kan i praksis blive overtalt til at benytte andre værktøjer til at måle resultatet af deres arbejde – men der kan være en del problemstillinger i forbindelse med evalueringen af arbejdet med udøvere, og disse problemstillinger kan ikke løses i det daglige arbejde med programmerne. Det er vigtigt at anerkende at solide evalueringsstudier er mere omfattende i deres koncept, og at der er brug for flere ressourcer for at kunne måle resultaterne. Det nødvendige arbejde og bestræbelser i sådanne evalueringsprojekter går meget længere end den daglige rutine indenfor måling af resultater, selvom det kan være en god begyndelse og basis at måle ens resultater, da dette kan betyde at ens DVPPs er ”klar til forskning” (se Gondolf, 2012, se figur 2). Som et resultat udtrykte behandlere nødvendigheden af ekstra finansiering, hvis de skulle være en del af evalueringsstudier.

Figur 2. Overlapning og specifikke sider af (a) evalueringsstudier, og (b) praksis til at overvåge resultaterne.



Generelt kan vi konkludere at der er et behov for evalueringsprojekter hvor flere lande er involveret – men der er et tydeligt behov for en harmonisering af metodologien. Det er

nødvendigt med DVPPs og støttecentrene for de udsatte, men for at øge muligheden for succes, bør fremtidige evalueringsstudier være planlagt som separate projekter med en finansiering der er stor nok til at dække alle ekstraomkostninger til de organisationer som vil være en del af projektet.

4.2. Hvor vi skal arbejde hen imod

Ideelt set vil den fragmenterede situation om nogle år, indenfor evalueringen af DVPPs være overstået, og der vil i stedet være en harmonisering af evalueringsmetodologien, så resultaterne fra de forskellige studier kan sammenlignes og arbejdes sammen. Ideen er at have en samarbejdende instans som fremmer, støtter og overvåger fremtidige DVPPS evalueringsprojekter, både når det kommer til udvikling, metoder, værktøjer, samt have en oversigt over hvilke spørgsmål der skal tages hensyn til og besvares. Dette vil ikke være et fuldt *DVPP-evalueringsforskningsprogram,* da det ikke er realistisk at en serie af studier vil ske i de næste par år, eller planlagt i omfattende maner, dette skyldes hovedsageligt manglende finansiering. Det er dog meget sandsynligt at der vil være flere evalueringsstudier, nogen vil finde sted på nationalt niveau (f.eks. som EU projekter) – og disse studier skal være sammenlignelige, og det skal være muligt at kunne kæde disse sammen på samme måde som et "puslespil".

Vi anbefaler, at man benytter sig af *Work-With-Perpetrators-European Network* (WWP-EN) som en samarbejdende instans, understøttet af *WAVE-Network*, af forskellige årsager: WWP-EN har taget føringen når det kommer til at sammenkæde DVPPs i EU og videre. Som en paraplyorganisation af DVPPs sammen med samarbejdende støttecentre for de udsatte, er der allerede taget skridt til at udarbejde standarder når det kommer til det aktuelle arbejde med udøvere af krænkende adfærd, og der vil sandsynligvis blive gjort endnu mere i fremtiden. Evaluering af dette arbejde er blot endnu et logisk skridt – det samme er gældende for **standarder** om hvordan dette **evalueringsarbejde** skal ske (se nedenfor).

Dette svarer til hvad Gondolf har kaldt "evalueringer ledet af behandlere" (2012, s. 81), dermed menes der evalueringsprojekter som er påbegyndt, udført i samarbejde med, eller som er inspireret af den praktiske side af arbejdet. Gondolf har påpeget at i mange tilfælde, har dette samarbejde mellem behandlere og forskere vist sig at være vanskeligt, idet hver part kun delvist forstår hinanden, hvilket dog også giver mulighed for forbedringer.

4.3. Standarder for evalueringsprojekter

Som vi har lært gennem Arbejdsgang 1 og 2, samt fra specifik litteratur, så er metodologiske standarder og et fælles sprogbrug meget vigtig, for at få en meningsfuld udvikling af fremtidige evalueringsprojekter. Det at evaluere arbejdet med udøvere og DVPPs, er generelt en meget specifik opgave, og procedurer fra andre områder (såsom evaluering af psykoterapi), kan ikke tilføjes i en form for "kopier & indsæt" metode. Nogle af de specifikke sider ved at evaluere DVPP er:

- *Koncept og sprogbrug:* Der har været mange dybdegående diskussioner om hvad DVPP skal stå for, med konsekvenser for udtryk og koncepter som f.eks. "succes". Der er

forskellige interessantgrupper som har udtrykt deres holdninger og meninger når det kommer til resultaterne af DVPP. Disse holdninger skal tages med i ens overvejelser, når fremtidige evalueringsstudier planlægges.

- *Informationskilder:* Information fra mændene, ses som ikke værende nok af de mange eksperter inden for området (se f.eks. Gondolf, 2012). Information fra partnere, politiet eller domstolene er en nødvendig del for at fuldføre de mange informationer. Information fra partnere ses som værende de mest troværdige og meningsfulde, og bør derfor være en central indikator. Dette kan dog også rejse nogle etiske spørgsmål som man skal huske og overveje (se Downes, Kelly & Westmarland, 2014).
- *Observationstider:* For at arbejde med de relative høje tal af afgange (både hvad angår udøvere, samt udsatte overlevende), samt muligheden for at vurdere de vedvarende ændringer som opnås i programmet, er det ikke nok med et enkelt før-design til at evaluere DVPPs. Hester et al. (2014) har foreslået et skema med observationspunkter som bør benyttes i denne sammenhæng. De har defineret fire forskellige observationspunkter: T0, til når en mand henvises til programmet; T1, når programmet påbegyndes; T2: under programmet; T3: i slutningen af programmet; T4: opfølgning, f.eks. seks måneder efter programmet er afsluttet. Dette skema sammen med krav til de informationer som indsamles på hvert observationstidspunkt (hvem skal spørges, hvilke værktøjer skal der bruges), kan bruges som en grundlæggende standard til evalueringer af DVPPs. Sammen med en tilstrækkelig beskrivelse af hvem mændene er på T0 (f.eks. demografisk eller psykologisk), samt hvordan de er blevet henvist til programmet, kan problemet med afgang blive studeret på en meget bedre måde, end det bliver gjort i dag.

4.4. Sådan fremmer vi standarderne

Når disse metodologiske standarder er blevet udviklet, skal de fremmes og benyttes i evalueringsstudier. I Arbejdsgang 3 blev der diskuteret forskellige strategier for at fremme disse standarder. De følgende muligheder er blevet identificeret:

- En arbejdsgruppe indenfor WWP-EN som er ansvarlige for problemstillinger indenfor evalueringer, kunne spille en vigtig rolle når det kom til at harmonisere en evalueringsmetodologi; ved at indsamle bedste praksis inden for DVPP forskning og ”styre” de mange serier af evalueringsstudier. Fremtidige evalueringsstudier skal kunne samarbejde med de programmer som allerede benyttes, på samme måde som det er sket tidligere. Det er derfor sandsynligt at evalueringsprojekter vil være baseret på WWP-ENs medlemsorganisationer, hvilket vil være en fordel når det kommer til evalueringsstandarder inden for WWP-EN.
- På samme måde vil de internationale netværk af støtteorganisationer til de udsatte, især *WAVE-Network*, spille en vigtig rolle. Når man evaluerer DVPPs er det vigtigt at få perspektivet fra de udsatte med – derfor vil sådanne projekter være i kontakt med, og samarbejde med, ikke blot WWP-EN medlemsorganisationerne, men også med WAVE medlemsorganisationerne. Eftersom WAVE og WWP-EN arbejder sammen, kan disse to Europæiske paraplyorganisationer have en betydelig indflydelse på harmoniseringen af evalueringsmetodologien, gennem deres medlemsorganisationer. Endvidere er der mulighed for at påbegynde (nationale og internationale) evalueringsstudier inden for WWP-EN. Når det kommer til EU-projekter som involverer flere lande, vil WWP-EN være den ideelle

Europæiske organisation til at koordinere de mange forslag inden for området af evalueringsforskning.

- Den næste milepæl vil være at fremme de metodologiske standarder, i et evalueringsprojekt som er finansieret af EU. Den konkrete metodologi kan blive udviklet som en del af dette fremtidige projekt, baseret på mulighederne som er vurderet i Arbejdsgang 3, hvor en mængde af mulige velegnede programmer fra forskellige lande er blevet identificeret under dataindsamlingen i Arbejdsgang 1 og af WWP-EN. Når det første projekt er blevet sat i gang til at evaluere DVPPs i flere lande, kan der opstartes flere projekter, som alle vil arbejde videre ud fra det første pilot-projekt.
- Fremmelse af forskningsmæssige praksisser i forbindelse med den værktøjssamling som er udviklet i Arbejdsgang 4. Selvom disse værktøjer er beregnet til at måle resultatet af den daglige praksis af WWP-programmerne, så er det et skridt i retningen af harmonisering af indsamling af data, og fremtidige studier i forskellige lande, kan arbejde videre på denne model. Når der kommer flere og flere programmer til, som benytter disse værktøjer, vil de indarbejde forskningslignende rutiner i deres daglige arbejde, som vil ligge den grundlæggende basis for et evalueringsprojekt. Disse målinger af resultater, vil forberede holdene bag disse programmer til ”forskning”.
- EU-programmer som harmoniserer forskning og praksis på tværs af Europa, er også en mulighed for at harmonisere standarder indenfor evalueringsforskning. Der er eksisterende EU-programmer som arbejder imod udviklingen af et Europæisk Forskningsområde, f.eks. COST. På dette niveau, vil netværk som WWP-EN og WAVE spille en førende rolle når det kommer til at fremme metodologiske standarder inden for evaluering, idet de vil hjælpe med at vælge programmer samt opmuntre deres medlemmer til at tage del i evalueringsstudier fra flere lande.

4.5. En uddybende tilgang til evaluering af DVPPS og vigtigheden af studier fra flere lande

Hvis vi begynder med ”puslespil” metaforen fra tidligere, kan forskellige former for evalueringsstudier, besvare forskellige spørgsmål – det vigtigste er at sikre, at disse studier kan blive sammenlignet og kædet sammen med hinanden. Dette kan sikres ved at benytte den samme metodologi (eller som minimum, så skal de basale procedure og værktøjer være sammenlignelige). Vi anbefaler at der kikkes på de forskellige evalueringsstudier, ud fra de kategorier som er nævnt nedenfor. (Der vil blive skrevet og offentliggjort mere uddybende artikler om dette, som en del af IMPACT projektet.)

4.5.1. Basale studier med single-kontekst

Dette er studier som evaluerer indvirkningen af DVPP, som del af en CCR. Sådanne studier udføres ude hvor selve arbejdet foregår, og ikke under laboratorie-lignende forhold, og som et resultat af dette, er der større mulighed for at der tilføres eksperimentelle designs, pga. tilfældige forskningsenheder (f.eks. mænd) til behandling af lidelser som oftest ikke er mulige, af metodologiske og etiske grunde. Klienter som har forladt programmet er blevet brugt som kontrolgrupper, eller andre grupper af mænd som har forladt programmet af andre grunde (f.eks. mænd som ikke blev lukket ind på programmet, eller som blev udelukket), disse sammenlignes med de klienter som har gennemført programmet.

Vi vil gerne anbefale tre elementer, som vi mener, bør være en del af en fremtidig evalueringsmetodologi, denne skal så efterfølgende tilrettes af *WWP-EN evalueringsarbejdsgruppe*: (a) observationer / tidspunkt; (b) indikatorer for succes; (c) design.

(a) Observationer / tidspunkt

Hester et al. (2014) har udarbejdet et skema gennem deres analyse af de Europæiske evalueringsstudier, for at beskrive "formen og størrelsen af et udsnit, på de forskellige faser af evalueringsprocessen" (p. 17). Dette skema bør bruges i evalueringsstudier af DVPPs, for at definere de forskellige tidspunkter for observationer. Dette skema består af fem punkter, kombineret med forslag til informationskilder på de forskellige observationstider:

- T0: "Før start". Det er vigtigt at vide hvilken slags mænd der henvises til programmerne (sociale demografiske data, psykologiske karaktertræk, misbrug, måden de er henvist på, osv.), samt om de ender med at være en del af programmet. Sammenligningen af udsnittene fra T0 og T1, sammen med baggrunden for klienter som forlader programmet igen, kan levere en meningsfuld information for fremtidige møder med de mænd som beslutter sig for ikke at være en del af programmet. Meget af denne information kan opnås gennem kontakt med de overlevende, visse informationer kan opnås gennem mændene, andre gennem behandlere. (Når det kommer til den etiske side af evaluering af programmer med overlevende, se Downes, Kelly & Westmarland, 2014.) Når det kommer til de ***overlevende,*** bør indikatorer for en succesfuld indblanding (se nedenfor) indsamles ved T0. Disse indikatorer vil være en central del af de målte resultater i T3 og T4.
- T1: "Start". Sammenligningen af udsnittet fra T0 og T1, vil give information om den første bølge af mænd som forlader programmet, sammenligningen af T1 og T2, senere hen T3, vil vise hvilke mænd der bliver på programmet, og hvilke der forlader det. Ændringer i deres liv samt information om kontekst vil være vigtige fra T0 til T3, for at kunne få en bedre forståelse af de forskellige grunde til at mænd forlader programmet. Når det kommer til ***mændene,*** vil fokus på dette tidspunkt være på faktorer som bør ændres gennem programmet (måling af resultater fra programmet), disse blev observeret ved T1 (f.eks. opførsel, holdning, motivation, risikofaktorer, osv.). Igen, kan de ***overlevende*** blive spurgt om deres indikatorer for en succesfuld indblanding.
- T2: "Under". Det anbefales at foretage gentagende målinger under hele forløbet. Udviklingen af indikatorer i løbet af forløbet kan give vigtig information, både angående mændene samt de overlevende. Man kan også adressere procedure relaterede spørgsmål, som efterfølgende kan forbindes med resultatet.
- T3: "Slutningen". Sammenligningen af behandlingsgruppen og kontrolgruppen ved T1 (før behandling) og T3 (efter behandling), er baggrunden for studierne, så man kan evaluere indvirkningen af behandlingen. Det er vigtigt at denne analyse inkluderer minimum to perspektiver: perspektivet fra de overlevende, i forbindelse med indikatorer for succes, samt perspektivet fra de mænd som har deltaget i programmet, samt dem fra kontrolgruppen, når det kommer til at måle resultatet af selve behandlingen. Information fra behandlere, og hvis muligt, offentlige instanser, kan indsamles ved T3.
- T4: "Opfølgning". Stabiliteten af de ændringer som er sket som en følge af programmet, skal vurderes med en opfølgning, dette bør ideelt ske seks måneder efter manden har afsluttet programmet. Igen er det vigtigt at stille spørgsmål til både de

overlevende og de mænd som har deltaget i programmet, man kan benytte behandlere og offentlige instanser som yderligere informationskilder.

En dropout analyse kan udføres på ethvert tidspunkt, både for mændene samt de overlevende ofte; dette kræver dog at den relevante information er indsamlet ved T0 (f.eks. demografisk, risiko, behov, osv.).

(b) Design

Kontrollerede Tilfældige Tests (RCT) ses som værende den bedste standard når det kommer til klinisk og psykologisk forskning. RCT er afhængig af en tilfældig tildeling af forskningsenheder (f.eks. mænd), til behandling af lidelser (f.eks. indblanding vs ingen indblanding). Der er en del problemer forbundet med denne procedure (f.eks. etiske problemer), og hvor RCTs er blevet benyttet, har de mødt megen kritik (se Gondolf, 2002, 2012). Selv i retssystemet diskuteres det ofte hvor svært det er at gennemføre RCTs (se f.eks. Dobash et al., 2000), selvom de bedste betingelser for at udføre en RCT, rent faktisk er i fængslerne (se nedenfor).

Figur 3. Design af en Tilfældig Kontrolleret Test (RCT)

Tilfældig		T1 (før)	(T2, under TM)	T3 (efter)	(T4, opfølgning)
tildeling	Behandlingsgruppe	O	X O	O	O
af individer					
til grupper	Kontrolgruppe	O	O	O	O

Som en konsekvens, har mange forskellige studier beregnet dropouts, ikke deltagere, eller mænd som er blevet ekskluderet fra programmet, som kontrolgrupper, hvor dem som har gennemført en DVPP er blevet defineret som en behandlingsgruppe. Dette betyder at behandlingsgrupper og kontrolgrupper ikke er sammenlignelige, hvilket udviser en mulig trussel til gyldigheden af sådanne studier (Shadish, Cook & Campbell, 2002). I hans berømte evalueringsstudie, har Gondolf (2002) fundet mindre effekter for dem der færdiggjorde programmet, sammenlignet med dropouts. For at kunne håndtere problemstillingen med kontrolgrupper som ikke er sammenlignelige, har Gondolf benyttet sig af tilbøjelighedsanalyser og variable værktøjer til at justere resultaterne. Disse procedurer førte til en del mere effektivitet. Princippet bag disse justeringer er følgende: forskellen mellem resultatet fra behandlingsgruppen og kontrolgruppen er statistisk set korrekt, ved at benytte forskellen mellem grupperne (dette skal være kendt ved begyndelsen af studiet) som en form for kontrol variabel. Som et resultat af dette, vil en effektivitetsstørrelse blive opnået som reflekterer hvad der kan forventes, under forudsætning af, at de to grupper var så sammenlignelige som muligt. Denne ændrede effektivitetsstørrelse ses som værende mere passende end den oprindelige, af eksperter indenfor statistikker (og den oprindelige størrelse, ses som værende ukorrekt).

Mange tidligere evalueringsstudier har været fokuseret på de mænd som deltog i programmet. Hvor de overlevende har været inkluderet som en informationskilde angående indvirkningen af programmet, har forskere oftest først været i kontakt med udøverne, og herefter prøvet at opnå kontakt med de respektive overlevende udsatte. Denne fremgangsmåde kan også benyttes i fremtidige studier, som Gondolf (2002) har fremvist i hans studie, har op til 60% af de oprindelige overlevende været kontaktet for en opfølgning.

Der er dog også alternative fremgangsmåde, som den følgende som er inspireret af MIRABEL projektet[1]. Fokus fra denne forskning kan blive flyttet fra mændene som deltager i programmet, til de overlevende i støttecentrene for udsatte. Hvis man tager i betragtning at der er geografiske regioner med støttecentre til de udsatte – men både med og uden DVPPs, kan den ekstra effekt ved en DVPP blive studeret når resultatet fra de overlevende måles i sådan en region eller by, og herefter sammenlignes til en region/by *uden* en DVPP (støttecentre til de udsatte er til stede i begge regioner). I denne opstilling vil de overlevende blive støttet og fulgt gennem en længere periode (T1-T4), det samme er gældende for deres partnere – som er henvist til en DVPP hvor der er en tilgængelige (behandlingsgruppe), eller som ikke kan henvises til en DVPP fordi der ikke er nogen (kontrolgruppe). Baseret på de overlevendes informationer om deres partnere, er det muligt at udføre en form for sammenlignelighed mellem behandlingsgruppen og kontrolgruppen ved T0. Selvom det virker nyttigt og meningsfuldt at inkludere mændenes perspektiv i evalueringen, og dermed forsætte som beskrevet ovenfor (”observationer/tidspunkt”), kunne man forestille sig blot at benytte sig af perspektivet fra de udsatte, kombineret med yderligere information fra behandlerne, i forbindelse med dropouts, afslappende holdning eller færdiggørelse af programmet. I sådanne tilfælde vil DVPP blive set som værende en ”sort boks”, som bruges til behandling af udøvere, i en bestemt mængde (afhængig af deres deltagelse). Behandlingsgruppen (overlevende hvis partnere er en del af programmet) vil herefter blive sammenlignet med kontrolgruppen (overlevende hvis partnere ikke er en del af programmet). Som en forudsætning skal kontekst i disse to regioner være sammenlignelige, så forskellen mellem behandlingsgruppen og kontrolgruppen kan ses som værende ikkeeksisterende i en DVPP. Selvom de detaljerede procedurer er mere komplekse[1] så er ideen med at studere indvirkningen af DVPP på de udsatte, som er en del af støtteprogrammer i andre regioner (med/uden DVPPs) meget lovende.

På lignende måde kan man udvikle andre design, f.eks. afbrudte tids-serier, hvor man følger de udsatte i en periode som svarer til forløbet af T0-T5, og hvor der på et tidspunkt vil blive oprettet DVPP i denne region. Hypotesen er at livskvaliteten af de udsatte vil blive forbedret, når deres mandlige partnere henvises til en DVPP på et systematisk basis. Tidsbaserede design er også velegnet til forskning af enkelte sager. I de tilfælde hvor man har et lille udsnit, skal man ikke undervurdere denne mulighed. Metodologi indenfor enkelt-sager er en lovende udvikling, også statistisk set, og kan bruges i kontekst når det kommer til evaluering af DVPPs. Det er fristende at se hvordan effekten af DVPPs kan studeres, baseret på data fra enkelt sager. Denne fremgangsmåde er især brugbar når forskellige personer modtager forskellige behandling (”doser”), afhængig af deres individuelle behov.

MIRABAL projektet har udtalt, at de vil benytte sig af en anden progressiv kvantitativ tilgang, til evaluering af programmer til udøvere, nemlig *overlevelsesanalyser.* Denne form for analyser kommer oprindeligt fra medicinsk forskning, og udregner forventninger til livet, men det bruges også inden for mange andre områder. Disse overlevelsesanalyser begynder med

tabeller over dødsfald, hvorfra der beregnes muligheder for overlevelse. Disse muligheder, som typisk vises som grafer over en vis tidsperiode, kan udarbejdes for forskellige undergrupper, f.eks. personer som modtager forskellige behandlinger. Der er også mulighed for at observere effekten af kovarianter. Vi anbefaler, at man benytter sig af denne innovative tilgang, i en fremtidig WWP-EN evalueringsarbejdsgruppe.

Resultaterne fra de basale evalueringsstudier som er beskrevet her, vil være udtalelser om den yderligere effekt som en DVPP har i en CCR (sammenlignet med en "CCR uden en DVPP").
Uanset hvilket design der vælges, så kan de forskellige studier kun sammenlignes, hvis T0-T4 skemaet bruges på en konstant måde (for at vide hvilken typer mænd der er med på de forskellige programmer, hvem der forlader programmet, og hvornår), samt hvis de samme resultater måles.

(c) Kriterier for succes, og måling af resultater

I et forudgående forskningsprojekt for MIRABEL projektet, har Westmarland, Kelly og Chalder-Mills (2010), foreslået de følgende "succeskriterier" for en DVPP:

- Forbedret forhold mellem mænd som deltager i programmet og deres partnere/tidligere partnere, respekt og forbedret kommunikation
- Udvidet hjælp til partnere/tidligere partnere, mulighed for at lade dem tale, deres mulighed for at tage egne valg, forbedret velvære
- Sikkerhed og frihed fra voldelige og krænkende adfærd over for kvinder og børn
- Sikker, positiv og delt forældrerolle
- Mænd skal blive bedre til at kende sig selv og andre, forstå påvirkningen som volden har haft på partnere og børn
- Sikrere, sundere barndom, børn skal føle at de bliver hørt, og at der er nogen der tænker på dem.

Disse kriterier blev brugt til at formulere de konkrete spørgsmål og enheder for MIRABEL studiet (se Westmarland & Kelly, 2012). Perspektivet fra de udsatte og deres børn er den primære informationskilde, i denne tilgang til at vurdere indvirkningen af DVPP. Denne information vil ideelt set blive suppleret med perspektivet fra de mænd som har deltaget i programmet, med resultater som nævnt ovenfor. Der kan tilføjes yderligere informationer fra behandlere og officielle instanser, som kan bruges til at besvare spørgsmål som går længere end en evaluering af resultatet i en DVPP. For at udarbejde en god praksis for fremtidige evalueringsstudier, anbefaler vi at en fremtidig WWP-EN evalueringsarbejdsgruppe tager disse indikatorer i betragtning.

4.5.2. Flersidede studier med multi-kontekst

Ved de *basale single-kontekst studier,* var der fokus på at bestemme påvirkningen af DVPPs, når forholdene er nogenlunde ens inden for de samme to områder. Nu vil vi kikke på en anden fremgangsmåde, hvor *sammenlignelige* DVPPs virker under *forskellige forhold.* Der er fremvist en struktur til hvordan denne opsætning kunne ske i figur 4. En DVPPs kontekst bestemmes af de relevante lokale komponenter som benyttes i forbindelse med indblanding

i tilfælde af vold i hjemmet (politiet, domstolene, beskyttelse af de udsatte, osv.), samt af de forhold som findes højere oppe i systemet, dvs. den finansielle situation på regionalt niveau, eller forskellige lovgivninger, kampagner og forebyggende forholdsregler som tages på et nationalt plan. Endvidere kan internationale og globale forhold have en indflydelse på systemets resultater.

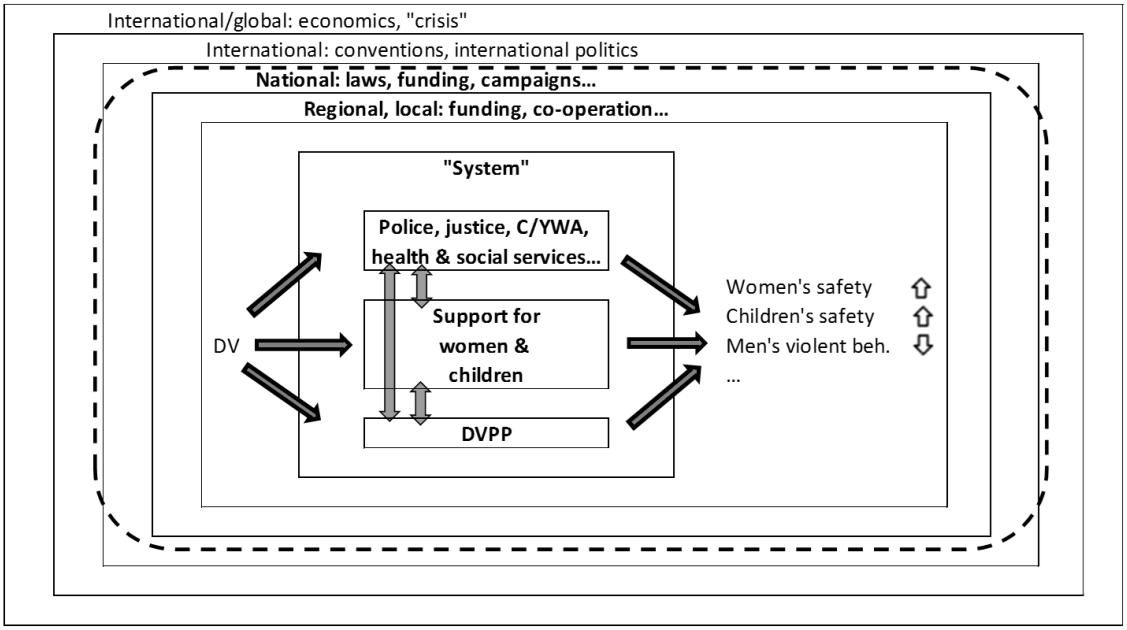
I de *basale single-kontekst studier,* har vi taget udgangspunkt i at kontekst er konstant, og at der er to områder (med eller uden DVPP) i et enkelt land, med den samme eller meget lignende lokal og regional kontekst. I *flersidede studier med multi-kontekst,* tager vi udgangspunkt i DVPPs som er meget ens, er håndteret under forskellige forhold, som det kan forventes i forskellige lande.
Resultaterne fra Arbejdsgang 1, har vist at lignende programmer eksisterer i forskellige lande – dette skyldes sandsynligvis at mange af programmerne er udviklet fra de samme grundlæggende principper, som f.eks. Duluth modellen eller CHANGE curriculum. Når to eller flere lande har meget lignende DVPPs som kan identificeres, ideelt set med den samme form for henvisninger (domstolene, selv-henvisende eller begge), så kan resultaterne fra disse systemer blive sammenlignet. I sådanne tilfælde kan vi antage, at der arbejdes konstant med DVPP, men at kontekst er forskelligt, hvilket fører til forskellige resultater. I sådan en opstilling, vil der blive foretaget en sammenligning mellem *betydningen af kontekst* og dens *påvirkning på resultatet af en DVPP,* disse resultater vil herefter blive studeret. Dette er i sådanne tilfælde at en fremgangsmåde med flere lande, giver mening.

Endnu engang skal man dog huske, at for at kunne kæde et sådan studie til andre evalueringer, er det nødvendigt at tilføje den samme grundlæggende metodologi i alle studier. Som et resultat af flersidede studier med multi-kontekst, kan vi forvente udtalelser om hvordan resultaterne fra en bestemt form for DVPP, influerer konteksten.

Hvis konteksten skal analysere dybere, uden at blive behandlet som blot endnu en "sort boks" (med en overvejende velkendt påvirkning, men med ukendte påvirkninger af de enkelt komponenter), så er udfordringen at *operationalisere konteksten*, for dermed bedre at kunne bestemme påvirkningen af komponenterne, på resultatet af en DVPP. Hvis man begynder med figur 4, er opgaven at omdanne konteksten ind til variabler til en kvantitativ analyse. De forskellige komponenter skal "oversættes" på en måde, så de kan passe sammen med kvantitative statistiske procedurer:

- På lokalt niveau: støtte til kvinder og børn, det arbejde som politiet udfører, retssystemet, børneforsorgen, sundhedsvæsnet og socialforsorgen;
- På regionalt og nationalt niveau: juridiske forbehold, lovforslag i forbindelse med at undgå vold, finansiering, generelle situationer med kønsdiskrimination.

Figur 4. En struktur for forskellige former for evalueringsstudier af DVPPs.



Bemærk. DV… Sager med vold i hjemmet; DVPP… Program for personer der udviser vold i hjemmet; C/YWA… Socialforsorgen. Sager med vold i hjemmet, møder et system der består af forskellige komponenter. Resultatet af systemets indblanding er afhængig af hver enkelt komponent, og deres indblandinger, samt de regionale, nationale og internationale forhold. Hvis man benytter perspektivet fra en DVPP, vil de andre komponenter af systemet, samt forhold på et højere plan (nationalt, internationalt), udgøre *konteksten* af en DVPP.

Sådan en operationalisering mangler i dag, og dens udvikling vil kræve betydelige ressourcer. Her vil vi blot give nogle få gode råd til om hvordan sådan en operationalisering kan foregå, samt forberede til fremtidige *flersidede studier med multi-kontekst*.

På *system niveauet,* er reaktioner fra domstolene blevet identificeret med en relevant indflydelse på resultaterne af DVPPs, i tilfælde af udøvere som er henvist af domstolene (f.eks. Gondolf, 2002). For at operationalisere denne indflydelse, skal man som udgangspunkt benytte perspektivet fra mændene som deltager i programmet, behandlere, de udsatte eller eksperter. Hvis vi benytter perspektivet af *mænd som er henvist til programmet af domstolene, som er vurderet af behandlerne,* kan de følgende indikatorer blive brugt:

A. Klienternes oplevelser og forventninger til domstolenes reaktioner

- "I tilfælde af manglende overholdelse, har der været (eller der kan forventes) en hurtig og konsekvent reaktion fra domstolen." (Værdi=1)
- "I tilfælde af manglende overholdelse, har der været (eller der kan forventes) en langsom og uregelmæssig reaktion fra domstolen" (0,5)[1]
- "I tilfælde af manglende overholdelse, har der ikke været nogen reaktion fra domstolen (og der kan ikke forventes nogen)." (0)

B. Usaglig mening i forbindelse med den offentliggjorte eller pålagte konsekvenser

[1]Or er et andet sæt af alternativer mellem 0 og 1.

- “Konsekvenser fra domstolen er meningsfulde / alvorlige i henhold til klienten.” (1)
- “Konsekvenser fra domstolen har en vis negativ indflydelse for klienten.” (0,5)
- “Konsekvenser fra domstolen er uden betydning for i henhold til klienten.” (0)

Denne information for alle sager A, B, A*B er nødvendig, fra alle relevante komponenter i et system, og fra forskellige perspektiver (mænd, behandlere, de udsatte). På denne måde kan megen information i forbindelse med konteksten blive indtastet i et data matrix som bruges til at udføre, f.eks. en multi-niveau analyse for at vurdere indflydelsen på variablerne og niveauerne af resultaterne af en DVPP. På et højere niveau kan eksperter vurdere reaktionerne fra domstolene på en mere *generel* måde, hvilket igen bliver indtastet som en variabel i et data matrix. De relevante indflydelser vil herefter blive besluttet empirisk i et sådan studie.

For at bestemme de relevante komponenter, kan udviklingen af et værktøj til at sammenligne kontekst, begynde med at sammenligne forskning ved juridiske og støttende midler som EIGE (2012). Der har også været tidligere projekter som har haft som mål at optegne og sammenligne kontekst i forbindelse med målinger af de udsatte i de forskellige Europæiske lande (f.eks. EUCPN, 2013). Dette arbejde viser hvilke juridiske og praktiske foranstaltninger der er i de forskellige lande.
Et andet eksempel for en mulig brugbar indikator er Gender Equality Index (GIA (ligestillingsindeks)), som er udviklet af EIGE (2012). Selv om vold ikke er en del af GIA grundet mangel på sammenlignelige data, så kan GIA benyttes som et overordnet indeks, når det kommer til landets placering i forbindelse med ligestilling, og de kan være en del af opstillingen som en uafhængig eller kontrolleret variabel.
En anden interessant tilgang er blevet foreslået af Hester og Lolley (2014), som har påpeget behovet for at adskille forskellige risikofaktorer for vold på forskellige niveauer, indenfor en økologisk struktur (socialt, offentligt, familie/nærhed, individuelle). Dette forslag er baseret på en interaktiv model af risikofaktorer som er blevet udviklet af Hagemann-White og kollegaerne (se Europæisk Kommission, 2010). Deres model kan bruges til at definere måder til at undgå voldelig adfærd i forhold. De relevante forebyggende komponenter kan blive operationaliseret, og brugt som variabler i en multi-niveau analyse.

4.5.3. Flersidede studier med single-kontekst

Fokus ved disse former for studier, er at sammenligne forskellige former for DVPPs indenfor en kontekst. Det er svært at bestemme hvad der fører til forskellige resultater, når *forskellige* programmer, studeres med *forskellig* kontekst – det anbefales dog, at sikre en så konstant kontekst som muligt, f.eks. ved at sammenligne to former for programmer i en region.

Denne tilgang vil ikke adressere de forskellige måder som man kan arbejde med udøvere på (f.eks. CBT eller psykodynamisk, eller med/uden ekstra ydelser, som f.eks. alkoholbehandling eller psykoterapi), men der kan også sammenlignes forskellige henvisningsmetoder, f.eks. CBT programmer i en region, hvor et program arbejder med klienter som kommer fra retssystemet, og et andet program arbejder med klienter som selv har henvendt sig. Fokus skal ikke blot være på hvilket program der er det bedste, men mere på, hvilken tilgang der virker bedst på de forskellige typer klienter – dette betyder, at det er nødvendigt at studere mændenes karaktertræk ved opstarten. Diskussionen om forskellige behandlinger til

forskellige udøvere, kan også adresseres gennem studier som sammenligner forskellige programmer (men indenfor et kontekst), endvidere kan man se på spørgsmålet om behandlingsintensiteten. Standardprogrammer kan bruges som udgangspunkt, hvorpå programmer med nye elementer kan sammenlignes.

Endnu engang er det vigtigt at der benyttes den samme form for metodologi i alle studierne, og i dette eksempel, vil den information som indsamles på T0 om mændene som deltager i programmet, være meget relevant. De flersidede studier med single-kontekst har den rigtige tilgang når det kommer til at lære mere om rammerne, metoderne, type af udøvere, osv. Som et resultat af dette, kan vi forvente konklusioner om forskellige effekter ved forskellige DVPPs, for forskellige typer udøvere.

4.5.4. Laboratorie-lignende studier

Endelig er der en form for evalueringsstudier hvor det ikke giver mening, eller ikke er muligt (f.eks. juridisk), at inkludere perspektivet fra de udsatte (f.eks. i forbindelse med programmer fra fængslerne i Spanien). På den anden side, er de eneste RCTs man kan finde, udarbejdet i lukkede institutioner som fængsler og psykiatriske hospitaler, dette kom frem under Arbejdsgang 2 (Hester et al., 2014). Fængsler giver en ramme hvor det ikke er muligt at observere indvirkningen af disse programmer med det samme, fordi de indsatte ikke bor sammen med deres partnere, og de generelt kun har begrænset adgang med dem. Ikke desto mindre er der en del programmer som benyttes i fængsler, som bruges til at forberede de indsatte på et liv efter fængslet, og de skal derfor ses som både en meningsfuld og vigtig indblanding.

Når det kommer til evaluering af DVPPs, kan man sammenligne programmerne i fængslerne med psykologiske eksperimenter som udføres i et laboratorium. I fængsler er det muligt at gennemføre RCTs. Især designet hvor der ventes (dvs. personer som endnu ikke modtager behandling, men er defineret som en kontrolgruppe som vil modtage behandling senere) er en mulig måde at gennemføre RCTs på i fængsler, uden de store etiske implikationer. Derfor kan studier som gennemføres i fængsler have som mål at forbedre programmerne, ved at fokusere på forskellige behandlingskomponenter, og ved at studere dem på en systematisk måde. Effektive komponenter kan implementeres i programmer uden for fængslet (og også afprøves og valideres som *flersidede studier med multi-kontekst*).
Som et resultat af laboratorie-lignende studier, kan differentierede udtalelser om effekten af behandlinger i fængsler (hovedsageligt baseret på mændenes perspektiv) forventes, samt udtalelser om effektiviteten af bestemte behandlingskomponenter. Sammenhængende studier med recidivisme kan også gennemføres – man skal dog huske, at recidivisme er afhængig af mange forskellige forhold og indflydelser, og ikke på hvad en mand lærer af et program i fængslet, så forventningerne til denne form for validering bør ikke være for høje.

5. Opsummering og de næste trin

Den oprindelige ide om et stort evalueringsstudie som involverer flere lande, som dækker over de Europæiske forskelligheder i fremgangsmåder når der arbejdes med udøvere, og som vil besvare en masse spørgsmål, skal tilrettes. Vi har fokuseret på arbejdet med udøvere i hjemmet og de samarbejde tilgange, af forskellige grunde, men også fordi eksperter inden

for forskning og praksis anbefalede os at gøre dette, gennem de forskellige workshops i dette projekt. Vi har beskrevet det basale for en metodologi inden for dette område, hvilket stadig skal udviklet mere detaljeret, før at vi kan bevæge os videre mod en mere systematisk og harmoniseret procedure, når det kommer til at evaluere DVPPs, både på et nationalt og internationalt niveau. Den nye internationale organisation WWP-EN kan og bør være en vigtig del af dette arbejde, fordi de kan udvikle og fremme standarderne for evaluering af DVPPs blandt deres medlemmer, og mange fremtidige evalueringsprojekter vil have brug for at samarbejde med disse medlemmer. Et meget vigtigt skridt som skal tages nu, er at fremskaffe en gruppe af interesserede forskere og behandlere inden for WWP-EN som vil have ansvaret for denne handling. Når de udvikler metoder og standarder, kan de begynde med resultaterne fra IMPACT projektet, samt især anbefalingerne i Arbejdsgang 3.

Der er visse detaljer når det kommer til evaluering af DVPPs – en af de vigtigste, er den viden at den mest troværdige information angående ændringer, kommer fra udøverens partner og børn. Som en konsekvens af dette, er det nødvendigt at arbejde sammen med organisationer som hjælper de udsatte, både på niveauet af enkeltstående centre, men også med paraplyorganisationer som WAVE. Det har vist sig at en systemtilgang er den mest passende, hvilket betyder at vi ikke skal arbejde hen imod at evaluere isolerede programmer for udøvere – men derimod evaluere hele systemet. Der er forskellige måder hvorpå dette kan gøres, og de er beskrevet i dette arbejdsdokument, ved at benytte nye udviklinger og eksempler på god praksis når det kommer til evaluering af DVPPs. Vi har foreslået en struktur for evalueringsstudier som betyder at man kan sammenligne og forbinde enkelte studier, samt indarbejde forskellige design til et meningsfuldt system, baseret på udviklingen af de andre Arbejdsgange i IMPACT projektet, tidligere studier samt nogle lovende udviklinger indenfor evaluering af DVPPs. Dette forslag vil forhåbentlig hjælpe til med at overvinde den spredte og problematiske situation som der eksisterer lige nu, når det kommer til sammenligning af evalueringer af DVPPs. Skridt for skridt, skulle vi være i stand til at opdage og fremvise de relevante effektstørrelser af en optimeret indblanding (”Hvem har brug for hvad”), for både udøvere og udsatte. Det at have troværdige effektstørrelser for DVPPs er også baggrunden for en meningsfuld og omkostningseffektiv analyse; indenfor de sociale og sundhedsmæssige instanser vil sådanne analyser normalt fremvise at de ressourcer som er tildelt til indblandinger, ikke blot omhandler etik – men de er også med til at spare på de offentlige budgetter. Det ses ofte at omkostningerne ved indblanding, er betydelig lavere, end sammenlignet med omkostningerne ved konsekvenserne af inaktivitet.
Yderligere vigtige aspekter og overvejelser vil blive offentliggjort i separate dokumenter, baseret på det arbejde som er foretaget i Arbejdsgang 3 i dette IMPACT projekt.

6. Litteratur

Alderson, S., Kelly, L. og Westmarland, N. (2013). Domestic violence perpetrator programmes and children and young people. London og Durham: London Metropolitan University og Durham University.

Alderson, S., Westmarland, N. & Kelly, L. (2012). The need for accountability to, and support for, children of men on domestic violence perpetrator programmes. Child Abuse Review, 3, 182-193.

Akoensi, T., Koehler, J. A., Lösel, F. & Humphreys, D. K. (2013). Domestic violence perpetrator programs in Europe, Part II: A systematic review of the state of evidence. International Journal of Offender Therapy and Comparative Criminology, 57 (10), 1206-1225.

Babbie, E. (2003). Badania społeczne w praktyce. Warszawa: Wydawnictwo Naukowe PWN.

Bamberger, M., Rugh, J. & Mabry, L. (2006). RealWorld evaluation. Working under budget, time, data, and political constraints. Thousand Oaks: Sage.

Cavanagh, K. & Lewis, R. (1996). Interviewing violent men: challenge or compromise? In K. Cavanagh & V.E. Cree (eds.), Working with men. Feminism and social work. London: Routledge.

Europarådet (2011). Council of Europe Convention on preventing and combating violence against women and domestic violence. Hentet på følgende link http://www.coe.int/t/dghl/standardsetting/convention-violence/convention/Convention%20210%20English.pdf

Cowburn, M. (2013). Men researching violent men: Epistemologies, ethics and emotions in qualitative research. In B. Pini & B. Pease (eds.), Men, masculinities and methodologies. New York: Palgrave Macmillan.

Dobash, R. E. & Dobash, R. P. (2000). Evaluating criminal justice interventions for domestic violence. Crime & Delinquency, 46 (2), 252-271.

Dobash, R. E., Dobash, P. R. (1998) Violent men and violent contexts. In R. E. Dobash & R. P. Dobash (eds.), Rethinking violence against women. Thousand Oaks, CA: Sage.

Dobash R. E., Dobash, P. R., Cavanagh, K. & Lewis, R. (2000). Changing violent men. Thousand Oaks, CA: Sage.

Downes, J., Kelly, L. og Westmarland, N. (2014) Ethics in violence and abuse research - a positive empowerment approach. Sociological Research Online, 19 (1) 2.

EIGE (2013). Gender Equality Index. Rapport. Vilnius: European Institute for Gender Equality. Hentet på http://eige.europa.eu/content/document/gender-equality-index-report

EUCPN (2013). Tackling domestic violence in the EU – policies & practices. In EUCPN Secretariat (eds.), EUCPN Toolbox Series, no. 4, European Crime Prevention Network: Brussels.

Europæisk Kommision (2010). Feasibility study to assess the possibilities, opportunities and needs to standardize national legislation on violence against women, violence against children and sexual orientation violence. Luxembourg: Publications Office of the European Union.

Geldschläger, H., Ginés, O., Nax, D. & Ponce, A. (2014). Outcome measurement in European perpetrator programmes: A survey. Hentet på følgende link http://wwp.webseiten.cc/fileadmin/WWP_Network/redakteure/IMPACT/Daphne%20III%20Impact%20Working%20Paper%201%20rev.pdf

Gondolf, E. W. (2002). Batterer intervention systems: Issues, outcomes, and recommendations. Thousand Oaks, CA: Sage.

Gottzen, L. (2013). Encountering violent men: Strange and familiar. In B. Pini & B. Pease (eds.), Men, masculinities and methodologies. New York: Palgrave Macmillan.

Hamilton, L., Koehler, J. A. & Lösel, F. (2013). Domestic violence perpetrator programs in Europe, Part I: A survey of current Practice. International Journal of Offender Therapy and Comparative Criminology, 57 (10), 1189-1205.

Hester, M. & Lilley, S.-J. (2014). Preventing violence against women: Article 12 of the Istanbul Convention. A collection of papers on the Council of Europe Convention on preventing and combating violence against women and domestic violence. Strasbourg: Europarådet.

Hester, M., Lilley. S.-J., O'Prey, L. & Budde, J. (2014). Overview and analysis of research studies evaluating European perpetrator programmes. Hentet på følgende link

http://wwp.webseiten.cc/fileadmin/WWP_Network/redakteure/IMPACT/Daphne%20III%20IMPACT%20Working%20Paper%202%20Overview%20and%20analysis%20of%20research%20studies%20evaluating%20European%20perpetrator%20programmes.pdf

Johnson, H. (1998). Rethinking survey research on violence against women. In R. E. Dobash & R. P. Dobash (eds.), Rethinking violence against women. Thousand Oaks, CA: Sage.

Kutrovátz, K. (2013) Work with Perpetrator – European Network national report Hungary. Hentet på følgende link
http://wwp.webseiten.cc/fileadmin/WWP_Network/redakteure/Reports_2013/National_report_Hungary.pdf

Patulny, R., Pini, B. (2013) Counting men: quantitative approaches to the study of men and masculinities. In B. Pini & B. Pease (eds.), Men, masculinities and methodologies. New York: Palgrave Macmillan.

Payarola, M. A. (2011) Description of a voluntary attendance program for abusive men. In C. Blazina & D. S. Shen-Miller (eds.), An international psychology of men. Theoretical advances, case studies, and clinical innovations. New York: Routledge.

Phillips, R., Kelly, L. & Westmarland, N. (2013) Domestic violence perpetrator programmes: A historical overview. London og Durham: London Metropolitan University og Durham University.

Price, B. J. & Rosenbaum, A. (2009). Batterer intervention programs: A report from the field. Violence and Victims, 24 (6),747-770.

Rossi, P. H., Lipsey, M. W. & Freeman, H. E. (2004). Evaluation. A systematic approach (7th ed.). Thousand Oaks: Sage.

Shaddish, W. R., Cook, T. D. & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin Company.

Silverman, D. (2013). Prowadzenie badań jakościowych. Warszawa: Wydawnictwo Naukowe PWN.

Stufflebeam, D. L. & Shinkfield, A. J. (2007). Evaluation theory, models, and applications. San Francisco: Jossey-Bass.

Westmarland, N. & Kelly, L. (2012) (samme med Kelly). Why extending measurements of 'success' in domestic violence perpetrator programmes matters for Social Work. British Journal of Social Work, 43(6), 1092-1110.

Westmarland, N., Kelly, L. & Chalder-Mills, J. (2010). What Counts as Success? London: Respect.

World Health Organization (2002). World report on violence and health. Summary. Hentet på følgende link

http://www.who.int/violence_injury_prevention/violence/world_report/en/summary_en.pdf

WWP European Network (2008). Guidelines to develop standards for programmes working with male perpetrators of domestic violence. Hentet på følgende link http://www.work-with-perpetrators.eu/index.php?id=86