



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

# Möglichkeiten für einrichtungs- und länderübergreifende europäische Evaluationsstudien zu Täterarbeits- Programmen bei häuslicher Gewalt

Working Paper 3 des Daphne III Projekts „IMPACT: Evaluation of European  
Perpetrator Programmes“

**Christian Scambor, Katarzyna Wojnicka & Elli Scambor**

**Deutsche Übersetzung: Elisabeth Müller**



## Index

1. Einleitung.....	3
2. Methoden.....	5
2.1 Literaturanalyse .....	5
2.2. Verwertung der Ergebnisse aus Workstream 1 .....	6
2.2.1. Wichtige Aspekte aus dem Workstream 1.....	7
2.2.2. Zusammenfassung.....	10
2.3. Verwertung der Resultate aus dem Workstream 2 .....	10
2.3.1. Wichtige Aspekte aus dem Workstream2 .....	11
2.3.2. Zusammenfassung.....	13
2.4. Interviews/Fokusgruppen mit ExpertInnen.....	13
3. Ergebnisse: Die Perspektive der ExpertInnen.....	15
3.1. Fragen.....	15
3.2. Möglichkeiten .....	16
3.3. Vorteile .....	17
3.4. Herausforderungen .....	18
3.5. Hindernisse .....	19
3.6. Was sich aus den ExpertInnen-Fokusgruppen und Diskussionen ergibt.....	20
4. Evaluierung von Täterarbeitsprogrammen bei häuslicher Gewalt: Wie es weitergehen kann.....	21
4.1. Ausgangspunkt: Der aktuelle Stand .....	21
4.2. In welche Richtung wir gehen sollen.....	23
4.3. Standards für Evaluierungsprojekte.....	23
4.4. Verbreitung der Standards .....	24
4.5. Ein umfassender Zugang zur Evaluierung von Täterarbeitsprogrammen bei häuslicher Gewalt und die Rolle länderübergreifender Studien.....	26
4.5.1. Grundlegende Studien in einem einzigen Kontext.....	26
4.5.2. Einrichtungs-und kontextübergreifende Studien.....	31
4.5.3. Einrichtungsübergreifende Studien in einem einzigen Kontext.....	34
4.5.4. Studien, die wie unter Laborbedingungen durchgeführt werden.....	34
5. Zusammenfassung und weitere Schritte.....	35
6. Literatur .....	37

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.

## Liste häufig verwendeter Abkürzungen

RKS...Randomisierte kontrollierte Studie

T0 ... Erhebungszeitpunkt 0: Bei Zuweisung; Aufnahmephase

T1 ... Erhebungszeitpunkt 1: „Prä“, bei Beginn der Intervention (z.B. Gruppenarbeit)

T2 ... Erhebungszeitpunkt 2: „Während“, während der Intervention (z.B. Gruppenarbeit)

T3 ... Erhebungszeitpunkt 3: „Post“, am Ende der Intervention (z.B. Gruppenarbeit)

T4 ... Erhebungszeitpunkt 4: „Follow-up“, eine Zeitspanne nach dem Ende der Intervention  
(z.B. 6 Monate nach dem Ende der Gruppenarbeit)

IMPACT ... Daphne III Projekt “Evaluation of European Perpetrator Programmes”

WAVE ... Women Against Violence in Europe (Internationale Organisation)

WWP – EN ... Work With Perpetrators – European Network (internationale Organisation)

## 1. Einleitung

Nach Auffassung von Fachleuten im Bereich der Täterarbeit ist eine Evaluierung von Täterarbeitsprogrammen von hoher Wichtigkeit für Fachkräfte und Programm-MitarbeiterInnen, um die Qualität ihrer Arbeit zu überprüfen und verbessern. Für EntscheidungsträgerInnen ist es sie ebenso wichtig, damit strategische Entscheidungen getroffen werden (Geldschläger et al., 2014; Hester, Lilley, O'Prey & Budde, 2014). Leider wurden bis dato nur wenige Evaluierungsprojekte Europäischer Programme in wissenschaftlichen Zeitschriften publiziert. Hester et al. (2014) haben jedoch zumindest 65 internationale Studien gefunden, die in Zeitschriften oder als graue Literatur veröffentlicht wurden. Die Studien machen allerdings einen fragmentarischen Eindruck. Die Mehrheit der Projekte wurde an einem einzigen Standort durchgeführt, nur wenige können als einrichtungsübergreifende Evaluierungsprojekte bezeichnet werden. Im Allgemeinen setzen sich die Berichte aus nationalen Studien zusammen, die auf unterschiedlichen Zugängen, unterschiedlichen Wirkmodellen, Methoden, Instrumenten etc. beruhen. Die Ergebnisse des Workstream 2 erwecken den Eindruck, dass das Gebiet der Europäischen Täterarbeitsprogramme breit gestreut und komplex ist und es daher schwierig ist, meta-analytische Rückschlüsse zu ziehen. Diese Situation führt dazu, dass die Europäischen Besonderheiten in den internationalen Diskussionen über Evaluierungsforschung von Täterarbeitsprogrammen kaum berücksichtigt werden. Jedoch gilt: „... existing evidence (e.g. from North American studies) cannot easily be generalised to a European context“ (Hester et al., 2014, p. 3).

Daher wurden die Hauptziele des von der Europäischen Kommission im Daphne III-Programm finanzierten Projektes „IMPACT: Evaluation of European Perpetrator Programmes“ folgendermaßen definiert:

- a) Die Wissenslücke bezüglich Evaluierungen von europäischen Täterarbeitsprogrammen zu füllen und
- b) Lösungen anzubieten, um die Ergebniskontrolle von Täterarbeit zu harmonisieren (als ersten Schritt in Richtung Harmonisierung von Evaluierungsmethoden).

Dies sollte folgendermaßen erreicht werden:

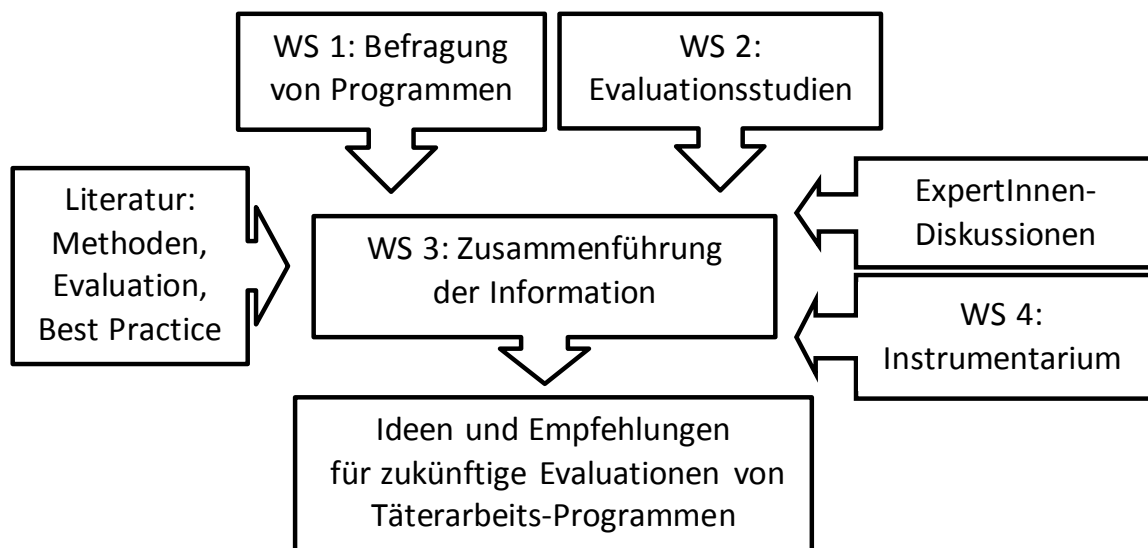
- Überblick und Analyse der aktuellen Praxis von Ergebniskontrolle europäischer Täterarbeitsprogramme zur Verfügung stellen (Workstream 1)
- Überblick und Analyse von Forschungsarbeiten, in welchen Täterarbeitsprogramme evaluiert werden, zur Verfügung stellen (Workstream 2)
- Möglichkeiten und Hindernisse für länderübergreifende Ergebnisforschung bestimmen (Workstream 3)
- Entwicklung eines Instrumentariums (ein sogenanntes „Toolkit“) und Good-Practice-Leitlinien für Ergebniskontrolle (Workstream 4)

Dieses Working Paper gibt einen Überblick über Methoden und Ergebnisse der Aktivitäten im Workstream 3. Die Ermittlung von Möglichkeiten und Hindernissen von einrichtungs- und länderübergreifender Ergebnisforschung basiert auf:

- Die Zwischen- und Endergebnisse des Workstream 1: Überblick über die bestehende Praxis der Ergebniskontrolle (Befragung)
- Die Zwischen- und Endergebnisse des Workstream 2: Überblick über europäische Evaluierungsstudien
- Durchsicht relevanter Literatur im Bereich Forschungsmethoden und Evaluierung
- Interviews und Diskussionen mit ExpertInnen und Fachleuten aus dem Bereich der Täterarbeit und kooperierender Einrichtungen.

Basierend auf den Ergebnissen der Befragung im Workstream 1 haben sich Fragen zu bestehenden vergleichbaren Methoden und Instrumenten verschiedener Organisationen zur Ergebniskontrolle- und Evaluierung ergeben. Es ergab sich auch die Frage, welche nationalen und institutionellen Rahmenbedingungen multinationale Evaluierungsprojekte fördern oder behindern. Aufbauend auf das durch die Sichtung der Forschung in Workstream 2 generierte Wissen wurden methodische Elemente gesammelt, die eine große Wahrscheinlichkeit aufweisen, dass ein Transfer in verschiedene nationale Gegebenheiten möglich ist. In diesem Sinne befasst sich Workstream 3 mit der Machbarkeit von Evaluierungsprojekten, die mehr als einen Standort und mehr als ein Land einbeziehen. Die Aktivitäten im Workstream 3 wurden mit dem Workstream 4 gekoppelt, um die Entwicklung des Ergebniskontrolle-Instrumentariums zu fördern, damit dieses Produkt dazu beiträgt, die Praktiken der Ergebnismessung von Programmen zu harmonisieren und so mitwirkt, die Voraussetzung für länderübergreifende Evaluierungsprojekte zu schaffen.

**Abbildung 1. Position und Funktion von Workstream 3 innerhalb des IMPACT-Projektes.**



Anmerkung. WS...Workstream.

In diesem Working Paper wird die Arbeit des Workstreams 3 zusammengefasst. Wir konzentrieren uns insbesondere auf den spezifischen Beitrag der ExpertInnen und auf unsere Schlussfolgerungen in Bezug auf zukünftige Evaluierungsprojekte im Bereich der Täterarbeit. Detailliertere Ausführungen der in diesem Working Paper behandelten Themen werden in zwei weiteren wissenschaftlichen Artikeln veröffentlicht.

Die AutorInnen bedanken sich bei allen PartnerInnen und assoziierten PartnerInnen des IMPACT Projektes, im Besonderen bei Heinrich Geldschläger und Marianne Hester und ihren Teams, die wesentliche Zwischenergebnisse für Workstream 3 zur Verfügung gestellt haben. Wir möchten uns auch bei allen ExpertInnen bedanken, die an Interviews und Diskussionen teilgenommen haben.

## 2. Methoden

Das Hauptziel des Workstream 3 war es, Möglichkeiten und Hindernisse zu bestimmen, bezogen auf einrichtungs- und länderübergreifende europäische Evaluierungsprojekte von Programmen mit männlichen Tätern, die Gewalt gegen Frauen und Kinder in ihren Beziehungen anwenden („Täterarbeitsprogramme bei häuslicher Gewalt“). Der Zugang, den wir wählten, um Evaluierungsmodelle zu ermitteln, die in unterschiedlichen europäischen Ländern Anwendung finden können, war der folgende:

- (1) Analyse bestehender Literatur zur Evaluierung von Täterarbeitsprogrammen (sowohl in europäischen als auch in außereuropäischen Kontexten) und die Analyse weiterer relevanter Literatur,
- (2) die Ergebnisse aus Workstream 1 und Workstream 2 gemeinsam mit zusätzlichen analytischen Schritten der in beiden Workstreams gesammelten Daten wurden verwendet
- (3) Interviews und Diskussionen mit internationalen ExpertInnen und Fachleuten, die mit Tätern häuslicher Gewalt arbeiten, wurden durchgeführt.

Es muss betont werden, dass der Schwerpunkt von Workstream 3 auf (a) *Ergebnisevaluation* (nicht Prozessevaluation oder anderen Aspekten zur Bewertung von Täterarbeitsprogrammen) und (b) auf *Evaluierungsforschung* lag, also über das Erheben von Ergebnissen in der täglichen Praxis der Täterarbeitsprogramme hinausgeht. (Ungeachtet dessen sind diese Aspekte miteinander verbunden und müssen oft gemeinsam betrachtet werden. Ergebnisse sind vom zugrundeliegenden Prozess abhängig; idealerweise sind Tätigkeiten der Ergebnismessung kompatibel mit den Methoden oder sogar Teil der Methoden, die in Evaluierungsprojekten verwendet werden.)

### 2.1 Literaturanalyse

Die Ergebnisse der Publikationen machten es möglich, einen internen Überblick über die häufigsten Evaluierungsmodelle, Instrumente, Strategien und Methoden zu erstellen und Fragen für die Interviews mit den ExpertInnen-Fokusgruppen und die ExpertInnendiskussionen zu formulieren. Wichtige Informationsquellen waren u.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); as well as two working papers from the IMPACT project: *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).

Eine weitere wichtige Informationsquelle für den Workstream 3 war methodologische Literatur über Evaluierung (z.B. Bamberger, Rugh & Mabry, 2006; Rossi, Lipsey & Freeman, 2004; Shaddish, Cook & Campbell, 2002; Stufflebeam & Shinkfield, 2007). Fachbegriffe und Glossare aus der Evaluierungsliteratur wurden den Projektgruppen zur Verfügung gestellt, damit während des gesamten Projektes eine einheitliche Terminologie verwendet werden konnte.

Gleichzeitig mit dem IMPACT Projekt wurde in Großbritannien ein wichtiges Evaluierungsprojekt mit dem Namen MIRABAL<sup>1</sup> umgesetzt. Dieses Projekt hat sich zum Ziel gesetzt, Fragen zu beantworten, die auch für das IMPACT Projekt relevant sind. Der Projektpartner *Respect* (Großbritannien) hatte im MIRABAL Projekt eine wichtige Rolle, wodurch der Transfer bedeutender Entwicklungen in beiden Projekten möglich war. Kontakt und Diskussionen unter den beiden Projekten waren eine wertvolle Informationsquelle (z.B. beim Workshop in Bristol, UK, im Jahr 2013, wo Nicole Westmarland die jüngsten Entwicklungen und die Methodologie des MIRABAL Projektes präsentierte). Eine wertvolle Informationsquelle waren auch die Publikationen in Verbindung mit dem MIRABAL Projekt (Alderson, Kelly & Westmarland, 2013; Alderson, Westmarland & Kelly, 2012; Downes, Kelly & Westmarland, 2014; Phillips, Kelly & Westmarland, 2013; Westmarland & Kelly, 2012; Westmarland, Kelly & Chalder-Mills, 2010). Das Ergebnis des MIRABAL Projektes wird Anfang 2015 präsentiert und wird für zukünftige einrichtungsübergreifende Evaluierungsprojekte sicherlich neue Impulse liefern. Es ist zu betonen, dass der US-amerikanische Experte Edward Gondolf (2012) das MIRABAL Projekt als Beispiel für einen neuen Trend in der Evaluierungsforschung sieht, den er „practitioner-led evaluations“ (S. 81) nennt. Diese Evaluierungen haben sich zum Ziel gesetzt, Defizite von manchmal ungeeigneten akademischen Zugängen zur Evaluierung von Täterarbeitsprogrammen auszugleichen.

## **2.2. Verwertung der Ergebnisse aus Workstream 1**

Workstream 3 hat sich auf Evaluierungsmodelle konzentriert, d.h. Designs, Instrumente und Abläufe von Evaluierungen. In Workstream 1 wurde eine Bestandsaufnahme durchgeführt, um Information über europäische Programme im Bereich der Täterarbeit bei häuslicher Gewalt und ihre Praktiken zur Ergebnismessung und Evaluierung ihrer Arbeit zu sammeln.

---

<sup>1</sup> <http://respect.uk.net/research/our-research-partnerships/mirabal-multi-site-evaluation-project/>  
<https://www.dur.ac.uk/criva/projectmirabal/>

Wir erwarteten, Information über vielversprechende Praktiken zu finden, die bereits vorhanden waren oder bei Evaluierungsprojekten verwendet wurden und einen gewissen Grad an Übertragbarkeit aufwiesen. Obwohl die AutorInnen erklären, dass „... the focus is not on scientific outcome research but on the day-to-day outcome monitoring perpetrator programmes regularly perform as part of their service delivery“ (Geldschläger et al., 2014, S. 4), haben wir erwartet, Hinweise über die Planung zukünftiger Evaluierungsstudien aus dem Workstream 1 zu bekommen. Wir wollten herausfinden, ob es einen gemeinsamen Nenner bei den Elementen gibt, die bereits vorhanden waren, und ob es Zugänge gibt, die in zukünftigen Studien angewandt werden könnten.

### *2.2.1. Wichtige Aspekte aus dem Workstream 1*

Bei der Bestandsaufnahme des Workstream 1 wurden Informationen aus 134 Programmen in 22 Ländern gewonnen. Diese Ergebnisse wurden in einem Working Paper (Geldschläger et al., 2014) veröffentlicht. Die wichtigsten Ergebnisse und ihre Folgen für den Workstream 3 werden unten zusammengefasst. Weitere Details sind bei Geldschläger et al. (2014) zu finden.

#### *Zuweisungen*

Die meisten Täterarbeitsprogramme arbeiten sowohl mit Selbstmeldern als auch mit Männern, die zugewiesen wurden. Ein Teil der Programme klammert manche Arten von Zuweisungen aus oder konzentriert sich auf eine bestimmte Art von Zuweisung (22% der Programme arbeiten ausschließlich mit Männern, denen Gerichte angeordnet hatten, am Programm teilzunehmen; 19% arbeiten überhaupt nicht mit von Gerichten zugewiesenen Männern; Geldschläger et al., 2014, S. 12). Viele Programme haben einen Schwerpunkt auf Männer gelegt, die von Gerichten zugewiesen wurden, oder auf Selbstmelder (oder auch von anderen Institutionen zugewiesene Männer) – die Mehrheit der Programme arbeitet jedoch mit verschiedenen Arten von Zuweisungen. Zukünftige einrichtungsübergreifende Evaluierungsstudien sollten diese Vielfalt von Zuweisungsarten berücksichtigen.

#### *Kooperation*

Fast alle Programme in der Bestandsaufnahme (98%) arbeiten mit Netzwerkpartnern zusammen. Die wichtigsten darunter: soziale Einrichtungen; Polizei; Opferschutz; Strafgerichte; Einrichtungen, die Unterstützung bei Alkohol- und Drogenmissbrauch anbieten. 75% der Programme geben an, dass sie Teil eines Kooperationsbündnisses gegen Gewalt sind und ein Teil der Programme, dass sie innerhalb ihrer Organisation Unterstützung für Opfer anbieten. Daher können wir schließen, dass ein „System-Zugang“ sehr ausgeprägt ist, bei den untersuchten Programmen in verschiedenen Ländern. Dass Art und Qualität der Netzwerke in den verschiedenen Ländern unterschiedlich sein können, sollte bei einrichtungsübergreifenden Evaluierungsdesigns berücksichtigt werden. Allerdings zeigt dieses Ergebnis auch, dass es möglich ist, sich auf diejenigen Programme zu konzentrieren, die schon Teil eines koordinierten Netzwerkes von Institutionen und Einrichtungen sind. Auch die Programme, die einen anderen Zugang haben, sind interessant – sie können Teil



eines Evaluierungsprogrammes sein und in einer zweiten Phase untersucht werden, weil dafür andere Vergleiche und Designs nötig sind.

### *Aufnahme*

Fast alle Programme (95%) wenden irgendeine Form von Aufnahmeverfahren an: von einem einzelnen Erstkontakt bis zu längeren individuellen Beratungs- und Diagnosephasen. Dies ist grundsätzlich für Evaluierungsprojekte nützlich: Viele Programme erfassen ganz selbstverständlich während des Erstkontaktes bestimmte Daten. Zwei Drittel der Programme verwenden bei der Aufnahme irgendeine Form von Verfahren – allerdings gibt es keine allgemein akzeptierten Verfahren. Eine Reihe von Fragebögen, Handbüchern und anderen Instrumenten werden genannt. Hier ist eine Harmonisierung nötig – es ist aber auch wichtig zu wissen, dass in den meisten Projekten *grundsätzlich* eine Verwendung von Verfahren Usus ist. Zukünftige Evaluierungsstudien können auf diese Kompetenzen aufbauen.

Die Mehrheit der Programme kann eine Form von Aufnahme- oder Ausschlusskriterien anwenden, die sich aus dem Aufnahmeverfahren ergeben (z.B. nennen drei Viertel „Alkohol- und Drogenmissbrauch“ und „psychische Krankheiten“ als ausschließende Kriterien). Das bedeutet, dass die Mehrheit der Programme Aussagen über bestimmtes Verhalten und psychologische Merkmale ihrer Klienten machen kann. Dies ist eine weitere wichtige Erkenntnis für einrichtungsübergreifende Evaluierungsstudien: Es scheint möglich zu sein, Verfahren zur Erfassung von Daten einzuführen, die vor dem Beginn des Programms stattfinden, damit später Unterschiede zwischen verschiedenen Gruppen von Teilnehmern und Abbrechern untersucht werden können.

### *Das Setting*

Ungefähr zwei Drittel der Programme wenden Gruppenarbeit an (die eventuell mit anderen Arten von Intervention erweitert wird); auch individuelle Beratung ist häufig, allerdings ist Gruppenarbeit die deutlich häufigste Intervention. Für einrichtungsübergreifende Evaluierungsstudien sind beide Zugänge (Gruppenarbeit; individuelle Arbeit) interessant. Andere Zugänge sind seltener (z.B. Paartherapie, Mediation); diese Zugänge könnten Teil bestimmter Evaluierungsformen sein, allerdings wären spezifische Evaluierungsmodelle nötig. In Bezug auf die „Dosis“ der Behandlung gibt es bei den Programmen große Unterschiede. Die Mehrheit der Programme wendet 14 bis 52 Einheiten in 14 bis 52 Wochen an. Diese Unterschiede sind für Evaluierungsstudien wichtig, in Bezug auf einen Dosis-Wirkungs-Zugang, was bedeutet, dass längere und kürzere Interventionen verglichen werden könnten.

### *Arbeitszugang*

In Bezug auf den Arbeitszugang der Programme, was die konkrete Arbeit mit Tätern betrifft, arbeiten die meisten Programme im Bereich der kognitiven Verhaltenstherapie und eines psycho-edukativen Zugangs. Zusammen mit einer Kombination aus kognitiver Verhaltenstherapie, psycho-edukativen und am Duluth-Modell orientierten Zugängen haben zwei Drittel bis drei Viertel so eine Grundlage. Diese Zugänge sind natürlich nicht identisch, haben aber manche Elemente gemeinsam, vor allem auf einer theoretischen Ebene (Lerntheorie). Für die U.S.A. nennt Gondolf (2012) zwei Hauptkategorien von Zugängen im

Bereich der Täterarbeit: kognitiv-behaviorale und psychodynamische Zugänge, beziehungsweise den „new psychology“-approach (Gondolf 2012), wobei jeder dieser beiden Zugänge in sich selbst viele Variationen aufweist. In Europa finden wir eine Vielzahl von Zugängen, die wiederum manchmal aus Elementen verschiedener Zugängen kombiniert sind.

### *Ergebnismessung*

81 % der Programme geben an, dass sie ihre Ergebnisse in irgendeiner Weise messen (S. 33). Die wichtigste Informationsquelle sind die Klienten selbst (bei 94% der Programme, die angeben, dass sie ihre Ergebnisse messen), TrainerInnen/GruppenleiterInnen (63%), und Ex-/Partnerinnen (60%), aber auch andere Institutionen (54%), Polizei- und Gerichtsberichte (43%) und Opferbetreuungseinrichtungen (32%) werden einbezogen. Follow-ups wurden von zwei Dritteln der Programme genannt.

Dieses Ergebnis zeigt, dass andere Informationsquellen als nur die Männer selbst in den Programmen häufig und üblich sind, worauf zukünftige Evaluierungsstudien aufbauen können. Die *Perspektive der Partnerin* wird als zentrale Informationsquelle in *Evaluierungsprojekten* angesehen. In den meisten Ländern gibt es Programme, die die Information der Ex-/Partnerin einbeziehen. Wie die Bestandsaufnahme der Programme zeigt, gibt es Länder, in denen dies üblich ist (z.B. Großbritannien) und andere, in denen dies nicht üblich ist (z.B. Frankreich, Schweiz). Es muss an dieser Stelle auch eine Verbindung zu *Programmen innerhalb des Strafrechtssystems* mancher Länder beachtet werden, nämlich: „... ‘mainly court-mandated’ programmes tend not to use information from partners and victim support services, as expected, because a lot of these programmes indicated that they don’t contact partners ... and don’t measure outcome with partners ...“ (Geldschläger et al. 2014, S. 46).

Allerdings verwenden Programme in 17 Ländern europaweit Informationen der Ex-/Partnerin bei ihren Ergebnismessungen. Daher schlussfolgern wir, dass es keine grundsätzlichen Probleme mit einer länderübergreifenden Studie geben sollte, die Partnerinnen in die Ergebnisbewertung der Programme einbezieht. Wo dies nicht möglich ist, werden die betreffenden Programme eine andere Rolle in einer zukünftigen Evaluierungsstrategie einnehmen.

In Bezug auf die Männer, die an den Programmen teilnahmen, waren die am häufigsten gemessenen Ergebnisse die folgenden: Gewaltfreiheit / Abnahme von Gewalt, Einstellungen und Überzeugungen, die Kommunikationsfähigkeiten des Klienten und Abnahme des Gewaltrisikos. Zwischen der Hälfte und zwei Dritteln der Programme erhoben auch psychologische Aspekte, Erziehungskompetenz und Lebensqualität.

Die wichtigsten Bereiche, zu denen die Ex-/Partnerinnen befragt wurden, waren: Gewalt Sicherheitsgefühl, Abnahme des Gewaltrisikos, Sicherheit der Kinder, Lebensqualität, Einstellungen des Mannes, die Kommunikationsfähigkeiten des Mannes, seine Erziehungskompetenz.

Alles in allem gibt es eine große Variabilität von Methoden, mit denen Ergebnisse von Täterarbeitsprogrammen gemessen werden. „One of the main conclusions of this survey is therefore that European perpetrator programmes are far from a consensus about standard methods of outcome measurement and a need for some harmonization seems quite

apparent.“ (Geldschläger et al., 2014, S. 56) Im Workstream 4 des IMPACT Projektes machen wir einen Schritt in diese Richtung.

### *2.2.2. Zusammenfassung*

Als Zusammenfassung der Bestandsaufnahme im Workstream 1 stellen wir fest, dass es für viele Länder und Programme keine grundsätzlichen Probleme bei einer umfangreichen Ergebnismessung gibt. Diese Messung bezieht die Perspektiven von Männern, die an Programmen teilnehmen, Partnerinnen und Kindern, TrainerInnen und Akten ein – es sind jedoch bei den meisten Programmen zusätzliche Ressourcen nötig, um an einem *Evaluierungsprojekt* teilzunehmen<sup>2</sup>. Fachleute scheinen für eine Evaluierung offen zu sein und verwenden diverse Verfahren zur Datenerhebung. Nichtsdestotrotz gehen Evaluierungsstudien über das Messen und Monitoring von Ergebnissen weit hinaus. Folglich sollten einrichtungsübergreifende Evaluierungsstudien als separate Projekte mit zusätzlicher finanzieller Förderung durchgeführt werden. Auf Grundlage der Erkenntnisse aus Workstream 1 erachten wir es als nicht machbar, einrichtungs- und länderübergreifende Evaluierungsstudien durchzuführen, die in Bezug auf Ergebnismessung und Erfassung von Information in der Aufnahmephase auf den gegenwärtigen Praktiken der Programme basieren. Bestehende Praktiken sind zu heterogen, um in einem einrichtungsübergreifenden Kontext nützlich zu sein. Die zusätzlichen Schritte in Evaluierungsstudien werden zusätzliche Ressourcen für praktisch alle Programme benötigen, damit sie sich an Evaluierungsprojekten beteiligen können.

### *2.3. Verwertung der Resultate aus dem Workstream 2*

Im Workstream 2 wurde der Schwerpunkt von der Praxis zur Forschung verlagert. Ziel war es, Erkenntnisse über die Zugänge, die in Evaluationsstudien über Täterarbeitsprogramme in Europa verwendet werden, zu erlangen. Das Forschungsteam erfasste sowohl Studien, die in wissenschaftlichen Zeitschriften veröffentlicht wurden, als auch in „grauer Literatur“, auf Englisch und in anderen europäischen Sprachen. Methoden, Designs, Ergebnismessung und weitere Aspekte der Studien wurden analysiert. Daher waren die Erkenntnisse aus dem Workstream 2 sehr wichtig für das Ziel des Workstream 3 (d.h. um zukünftigen Evaluierungsstudien Empfehlungen in Bezug auf Evaluierungsmodelle zu geben).

---

<sup>2</sup> Dies trifft auch auf Ergebnismessung von Täterarbeitsprogrammen zu: Das Hauptproblem bzw. -Hindernis für viele Programme sind beschränkte Zeit und beschränkte Ressourcen. Ein geringerer Anteil der Programme (ca. ein Drittel) sehen „fehlende Methodologie“ und „rechtlicher/institutioneller Kontext“ als hinderlich für die Ergebnismessung (rechtliche Hindernisse wurden schon oben genannt, betreffend Programme im Rahmen des Strafrechtssystems mancher Länder). Allerdings sind fast alle Programme in der Bestandsaufnahme daran interessiert, ihre Ergebnismessung durch Evaluierung und internationalen Austausch zu verbessern. Daher wünschen sich viele Fachleute mehr Arbeitskapazität, wirtschaftliche Ressourcen, Methoden, Ressourcen für die Datenanalyse, Richtlinien/Empfehlungen und Training/Beratung.

### 2.3.1. Wichtige Aspekte aus dem Workstream 2

#### *Häufigkeit nach Ländern*

Hester et al. (2014) beschreiben ihren Zugang zur Auffindung vorhandener Studien sehr genau. Vier Meta-Analysen und Zusammenstellungen von Studien wurden gefunden, die auf insgesamt 15 Studien basieren. Interessanterweise sind die meisten dieser Studien aus Großbritannien (7) und Spanien (5) Diese beiden Länder waren auch unter den Ländern, die in Workstream 1 als die Länder mit den meisten Programmen erfasst wurden.

Gemeinsam mit weiteren Schritten zur Erfassung von Evaluierungsstudien, die sowohl veröffentlicht als auch unveröffentlicht waren (oder als graue Literatur veröffentlicht sind), konnten insgesamt 65 Studien für die weitere Analyse im Workstream 2 verwendet werden. Wiederum kommen viele dieser Studien aus Spanien und Großbritannien und auch aus zentraleuropäischen Ländern.<sup>3</sup> Wir stellen fest, dass sich die Häufigkeit der *Programme* (Workstream 1) mit der Häufigkeit von *Evaluierungsstudien* (Workstream 2) deckt. Das ergibt Sinn: Länder, die mehr (sichtbare) Täterarbeitsprogramme haben, führen mehr Evaluierungsstudien durch und veröffentlichen diese Studien auch.

#### *Probleme in Bezug auf die Studien*

Bei der Analyse der Evaluierungsstudien ergab sich eine Reihe von Problemen. Einige der wichtigsten werden hier aufgelistet:

- *Design*: Es wurden kaum randomisierte kontrollierte Studien gefunden. Unter n=58 Studien aus fünf europäischen Regionen, die sich mit *Ergebnisevaluation* befassen (sie konzentrierten sich nur auf Ergebnisevaluation – oder auf Ergebnis und Prozess zusammen), waren nur zwei randomisierte kontrollierte Studien, und diese beiden Studien wurden in einem Gefängnis und einer psychiatrischen Klinik durchgeführt.
- *Kontrollgruppen*: Vielen Studien fehlt es an einer Kontrollgruppe oder sie haben nur kleine Kontrollgruppen. Diese Tatsache verweist auf die wesentlichste Einschränkung in den verfügbaren Studien, da Kontrollgruppen (oder äquivalente Elemente von Kontrolle, die Rückschlüsse auf ursächliche Zusammenhänge erlauben) essentiell für die interne Validität jeder Studie sind.
- *Terminologie*: In den Studien wurden viele unterschiedliche Definitionen zentraler Begriffe festgestellt (z.B. zu den Begriffen „Beginn“, „Abschluss“). Es gibt keine einheitliche Terminologie.
- *Zeitpunkte*: Es gab Unterschiede in den Zugängen und Methodologien in Bezug auf die Zeitpunkte, an denen Daten erfasst wurden. Hester et al. (2014) haben für ihre Analyse ein Schema erstellt, dessen Verwendung für ein gemeinsames Verständnis gefördert werden sollte. Das Schema besteht aus fünf Zeitpunkten, zu denen bei einer Evaluierung Information erfasst werden soll (T0: Vor dem Beginn der

---

<sup>3</sup> Für ihre Analyse verwendeten Hester et al. (2014) eine Klassifizierung der Länder in fünf Regionen, in denen Evaluierungsstudien erfasst wurden: Westen (Großbritannien, Irland), Süden (Spanien, Portugal), Norden (Schweden, Dänemark, Finnland, Island), Osten (Kroatien) und Zentral (Deutschland, Österreich, die Schweiz und die Niederlande).

Intervention [Prä]; T1: zu Beginn der Intervention; T2: Während der Intervention; T3: bei Abschluss der Intervention; T4: Follow-up). Dieses Schema wird weiter unten genauer beschrieben.

- *Ausfall*: Ausfallraten waren unklar oder es war unklar, in welchen Phasen Ausfälle auftraten. Informationen zu Stichprobengrößen und Ausfällen in *jedem* Stadium des Interventionsprozesses fehlten in den meisten Fällen. In Bezug auf Stichproben und Stichprobengrößen variieren die vorhandenen Informationen stark. Dieser Mangel ergibt sich aus fehlenden Standards und fehlender Standard-Terminologie für Evaluierungsstudien zu Täterarbeitsprogrammen.
- *Zuweisung*: Information über die Art der Zuweisung oder Programmeinstiege (Selbstmelder oder vom Gericht angeordnete Teilnahme oder andere Arten der Zuweisung) ist entscheidend. Es wird auch Information über die Gründe von Ausfällen benötigt, sowie sozio-demographische und andere Informationen von allen Teilstichproben (alle zugewiesenen Klienten, bevor das Training startet; Klienten am Beginn des Trainings; Ausfälle, Abschießer). In vielen Fällen fehlt diese Information.

### *Ergebnismessung und Auswirkung*

In allen Regionen wurden Daten verwendet, die von den Tätern selbst, aus Akten des Strafrechtswesens und von den Opfern stammten. Dies trifft jedoch nicht auf den Süden zu, wo Daten von Partnerinnen in den vorhandenen Studien nicht aufschienen. Wie wir aus der Analyse im Workstream 1 wissen, gibt es aber einzelne Programme in der südlichen Region, die die Opferperspektive einschließen. Allerdings scheint es einen Unterschied zwischen der Praxis und den rechtlichen Möglichkeiten von Programmen in Gefängnissen und inter-institutionellen Programmen zu geben. Programme im Rahmen des Strafrechtssystems integrieren die Perspektive der Partnerinnen nicht, werden aber öfter evaluiert als inter-institutionelle Programme außerhalb der Gefängnisse. Außerdem verwenden die Evaluationen im Strafrechtssystem viele validierte/standardisierte psychometrische Instrumente, und sie konzentrieren sich auf psychologische Aspekte und Veränderungen ihrer männlichen Täter. Dies spiegelt einen starken psychologischen Zugang zur Evaluierung von Täterarbeitsprogrammen im Strafrechtssystem der südlichen Region wider. Die meisten dieser Studien kommen aus Spanien.

Die umfassendsten Ergebnismessungen wurden in der westlichen Region gefunden, „... including change in attitudes towards women and violence against women, official and self-reported recidivism, repeat victimisation, the quality of life, safety and well-being of the women/victims and their children, levels of parenting stress and drug / alcohol use of the perpetrator.“ (Hester et al., 2014, S.26) Verfahren der Risikoeinschätzung finden sich in Studien der westlichen und südlichen Region.

Im Allgemeinen variieren die erforschten Bereiche und Verfahren je nach Region (und Studie) stark. Dieses Fehlen einer einheitlichen Methodologie erschwert den Vergleich von Studien miteinander oder macht ihn sogar unmöglich.

Wenn nun also auf Basis der verfügbaren Evaluierungsstudien eine Gesamtaussage zur Wirksamkeit von Täterarbeitsprogrammen gemacht werden soll, kann nur ein sehr allgemeines Bild gezeichnet werden: Es gibt entweder einige statistisch signifikante Erkenntnisse, die in die Richtung einer positiven Veränderung zeigen, oder es gibt vielversprechende Ergebnisse, die aber statistisch nicht signifikant sind. Dieses Muster

wurde bei einer Vielzahl von Bereichen gefunden (Verhaltensänderung; Veränderung der Einstellungen; psychologische Veränderungen; Veränderungen in Bezug auf Risikofaktoren; Motivation und Mitarbeit). Allerdings gibt es Einschränkungen der Ergebnisse in Bezug auf die Wirkungszusammenhänge, nämlich die häufig kleinen Stichprobengrößen, hohe Ausfallraten (oder fehlende Information über Ausfälle) und zu große Abhängigkeit von der Information, die die Klienten selbst liefern (betrifft die südliche Region).

### 2.3.2. Zusammenfassung

Trotz aller Probleme, die bei den Studien festgestellt wurden, sprechen die Ergebnisse dafür, dass es möglich ist, aus allen Regionen Daten zu erhalten und diese auch vergleichen zu können. Es gibt zwar unterschiedliche Zugänge zur Evaluierung, aber es scheint möglich, bestimmte Verfahren anzuwenden und bestimmte Informationsquellen in den Ländern einzuführen (mit einigen Ausnahmen, wie z.B. Programme in Gefängnissen, die eine andere Rolle in zukünftiger Evaluierungsstrategie einnehmen müssen).

Das aktuelle Hauptproblem scheint die uneinheitliche Methodologie in Bezug auf die Evaluierung von Täterarbeitsprogrammen zu sein, was den Vergleich in einigen Aspekten (Design und Kontrollgruppen, Terminologie, Verfahren, Informationsquellen) schwierig macht. Eine Harmonisierung der Evaluationsmethodologie ist nötig, um die Fragmentarisierung von Evaluierungsforschung zu Täterarbeitsprogrammen auszugleichen. Laut der AutorInnen des Working Papers von Workstream 2 scheint dies möglich:

„There were some interesting differences and similarities between and within regions and it would be possible in the main to take elements from different approaches in order to start developing a robust evaluation methodology.” (Hester et al., 2014, S. 39)

### 2.4. Interviews/Fokusgruppen mit ExpertInnen

Bei Evaluierungsstudien zu Täterarbeitsprogrammen gibt es eine Reihe von Interessensgruppen mit ihren eigenen Interessen und Perspektiven, die zum Teil unterschiedlich sind, sich aber auch überschneiden können:

- *Opfer häuslicher Gewalt:* Die Partnerinnen und Kinder gewalttätiger Männer werden wissen wollen, ob ein Programm wirksam ist, das gewalttätige Verhalten ihrer Partner/Väter dadurch aufhören wird und sich somit ihre Sicherheit erhöht.
- *Fachleute im Bereich Opferbetreuung:* Fachleute in Organisationen zur Opferbetreuung möchten wissen, ob Programme wirksam sind bzw. ob sie wirksam genug sind, um die Verwendung zusätzlicher finanzieller Mittel zu rechtfertigen, die ansonsten für verstärkte Opferbetreuung verwendet würden.
- *Fachleute in der Interventionskette:* Polizei, Staatsanwaltschaft, Gerichte, BewährungshelferInnen und andere Interessensgruppen sind ebenfalls daran interessiert, ob Täterarbeitsprogramme eine bestimmte Wirkung haben (zusätzlich zu bereits bestehenden Interventionen).
- *Fachleute in Täterarbeitsprogrammen:* Diese Fachleute werden daran interessiert sein, was ihre Arbeit bewirkt bzw. genauer gesagt, ob ihre Arbeit effektiv ist und sich

Gewalt dadurch verringert. Sollte dies der Fall sein, werden sie wissen wollen, wie sie ihre Arbeit verbessern können, um noch günstigere Ergebnisse zu erzielen.

- *ProgrammleiterInnen von Täterarbeitsprogrammen*: Sie werden sich sowohl für die Wirksamkeit der Programme als auch für die Qualitätsverbesserung der Programme interessieren, d.h. es ist in ihrem Interesse, Wirksamkeit nachzuweisen und Ergebnisse zu verbessern.
- *Die Öffentlichkeit*: Die Öffentlichkeit kann allgemein dran interessiert sein, dass sich häusliche Gewalt in der Gesellschaft verringert und sich die Sicherheit erhöht. Die öffentliche Diskussion zum Thema findet oft in den Medien statt, die ihre eigene Logik und ihre eigenen Interessen haben. Qualitätsmedien können Fragen stellen, die denen anderer Interessensgruppen ähnlich sind, andere Segmente in den Medien arbeiten dagegen mit schockierenden, auffallenden und außergewöhnlichen Inhalten; von dieser Seite hört man oft den Ruf nach härterer Bestrafung der Täter und strengeren Gesetzen, es gibt aber auch dort teilweise Interesse an der Wirksamkeit der Alternativen, wie sie z.B. Täterarbeitsprogramme darstellen.
- *GeldgeberInnen, PolitikerInnen*: Interessensgruppen dieser Kategorie möchten der Öffentlichkeit gegenüber ihre Bemühungen gegen Gewalt nachweisen können, aber auch sicher gehen, dass ihre Geldmittel sinnvoll angelegt werden und ihre vordefinierten Ziele erfüllt werden; gleichzeitig möchten sie die Verwendung ihrer Ressourcen optimieren, d.h. dass sie Antworten auf die Frage nach der Kosteneffizienz wünschen.

In den *Interviews unter ExpertInnenfokusgruppen* und bei den *ExpertInnendiskussionen* (vgl. unten) wurden sowohl die unterschiedlichen Interessen als auch die Überschneidungen dieser Interessen thematisiert.

Die **Interviews/Fokusgruppen mit ExpertInnen** (eine Methode der Sozialforschung, die ExpertInneninterviews mit Fokusgruppeninterviews kombiniert; vgl. Babbie, 2003; Silverman, 2013) wurden während des *Projektworkshops A* im Oktober 2013 an der University of Bristol, Großbritannien, durchgeführt. Vier 45-minütige Fokusgruppeninterviews wurden durchgeführt, mit 18 ExpertInnen (ForscherInnen und Fachleuten, die mit Tätern arbeiten) aus 11 europäischen Ländern (Österreich, Tschechien, Dänemark, Finnland, Deutschland, Irland, Polen, Portugal, Norwegen, Spanien, Großbritannien). In jeder Gruppe stellte der/die ModeratorIn (ein Mitglied des Projektteams) drei wesentliche Fragen:

- Welche fördernden und welche hinderlichen Umstände in Bezug auf die Machbarkeit eines einrichtungs- und länderübergreifenden Designs sehen Sie, auf der nationalen und lokalen Ebene?
- Welche fördernden und welche hinderlichen Umstände in Bezug auf die Machbarkeit eines einrichtungs- und länderübergreifenden Designs sehen Sie, auf der internationalen Ebene?
- Welche Möglichkeiten und welche Hindernisse sehen Sie in Bezug auf folgende Idee: „Ein Instrumentarium für Fachleute (in Täterarbeitsprogrammen) zur Verfügung stellen, vergleichbare Daten in verschiedenen Ländern sammeln; diese Daten zusammenspielen und zentral analysieren“?

Nach den Interviews erstellten die GruppenleiterInnen interne Berichte, die als Basis für die Präsentation im *Projektworkshop B* und auch zur weiteren Analyse dienten (vgl. unten).

Die **ExpertInnendiskussion** fand während des Workshops B statt, der im April 2014 in Kopenhagen, Dänemark abgehalten wurde. 41 Fachleute, ForscherInnen und EntscheidungsträgerInnen aus Österreich, Bulgarien, Tschechien, Kroatien, Dänemark, Estland, Finnland, Deutschland, Irland, Italien, Litauen, den Niederlanden, Russland, Spanien, Schweden, der Schweiz und Großbritannien diskutierten die Themen und Fragen, die sich aus der Präsentation *Evaluierungsmöglichkeiten mit einem einrichtungs- und länderübergreifenden Zugang* ergaben:

- Welche Möglichkeiten und Hindernisse erkenne ich in meiner Organisation?
- Welche Fragen sollten bei Evaluierungsstudien gestellt werden:
  - „Funktionieren“ Täterarbeits-Programme, wirken sie?
  - Bis zu welchem Grad / wie effektiv sind sie?
  - Was wirkt, welche sind die Wirkfaktoren?
  - Spezifische Fragen (Wer wird rückfällig, wer braucht welche Intervention usw.)

Die Ergebnisse der Diskussionen der vier Gruppen wurden gesammelt und dienten als Grundlage einer weiteren Analyse (vgl. unten). Weiteres Feedback zum Konzept einrichtungs- und länderübergreifender Evaluierungsstudien von Täterarbeitsprogrammen wurden bei der Schlusskonferenz des Projektes in Barcelona, Spanien gesammelt (Oktober, 2014).

### 3. Ergebnisse: Die Perspektive der ExpertInnen

Die Ergebnisse der Interviews/Fokusgruppen mit ExpertInnen und der ExpertInnendiskussionen wurden in einer Sammlung von Themen zusammengefasst, die eine Informationsquelle dafür darstellen, wie mit der Entwicklung einer Methodologie für die Evaluierung von Täterarbeitsprogrammen weiter vorgegangen werden soll. Diese Themen wurden in fünf Hauptgruppen eingeteilt: Fragen, Möglichkeiten, Vorteile, Herausforderungen und Hindernisse.

#### 3.1. Fragen

Eine der wichtigsten Fragen, die sich während der Interviews und Diskussionen ergab, bezog sich auf das grundsätzliche Verständnis von „Täterarbeitsprogrammen/Täterarbeitsprogrammen“. Laut der ExpertInnen aus verschiedenen Ländern gibt es unterschiedliche Interpretationen solcher Programme: Abgesehen von typischen Programmen (d.h. regelmäßige Gruppentreffen, die von ausgebildeten TrainerInnen geleitet werden, strukturiert und koordiniert durchgeführt sowie evaluiert werden) wurden die folgenden Interventionen genannt, die direkte oder indirekte Teile von Programmen für Männer sein können, die in ihren nahen Beziehungen gewalttätig werden: Telefonhelplines; individuelle Settings; informelle Treffen, ehrenamtliche Treffen,



kirchliche Treffen, Treffen der Anonymen Alkoholiker usw. Daher wäre der erste Schritt am Weg zu einer länderübergreifenden europäischen Evaluierungsforschung zu Täterarbeit eine gemeinsame aber relativ offene **Definition von „Täterarbeits-Programm“**.

Eine weitere wichtige Frage betrifft die **Ähnlichkeit von Programmen**. Die meisten ExpertInnen teilen die Meinung, dass nur Ergebnisse ähnlicher Programme (gleicher Zugang, ähnliche Arbeitsweise, Dauer, Gruppengröße usw.) miteinander verglichen werden sollten. Diese Annahme hat allerdings Zweifel aufkommen lassen, in Bezug auf die (unzureichende) Anzahl von europäischen Programmen, die einander ähnlich genug sind, um sie miteinander zu vergleichen zu können.

Nach der Frage nach der Ähnlichkeit wurde die Frage nach der **Art von Programmen** gestellt, die untersucht werden sollen. Manche ExpertInnen betonen, dass sich zukünftige Studien auf nur eine Art von Programm konzentrieren sollen: entweder angeordnete Programme oder Programme für Freiwillige. Andere ExpertInnen wiederum schlagen vor, beide Zugänge zu erforschen oder sich auf „gemischte“ Programme zu konzentrieren (d.h. mit beiden Zuweisungswegen im gleichen Programm).

Schlussendlich besprachen die ExpertInnen das Thema **Forschungsverfahren in** einer zukünftigen Studie. Konkrete Fragen überwogen, so wie die Frage, wer forschen soll, wer Daten erheben soll usw. Die Vor- und Nachteile der verschiedenen Möglichkeiten wurden diskutiert, z.B. ob unabhängige, externe ForscherInnen die Erhebung und die analytischen Schritte durchführen sollen, oder aber die Fachleute und Personen, die direkt mit den Tätern arbeiten. Motivation und ernsthaftes Interesse an der Evaluierung von Täterarbeitsprogrammen könnten die Vorteile sein, wenn Fachleute und TäterarbeiterInnen diese Tätigkeiten selbst durchführen. Ein Nachteil wäre allerdings der mögliche Interessenskonflikt. Den ExpertInnen zufolge haben beide Vorgangsweisen ihre Vor- und Nachteile (unterschiedlicher Zugang zu den Daten und Stichproben; Kosten; unterschiedliche Forschungskompetenzen; das Thema „Objektivität“; Zeit usw.). Die Entscheidung für ein bestimmtes Verfahren soll also gut durchdacht sein.

### 3.2. Möglichkeiten

Viele ExpertInnen bekundeten ihr Interesse daran, eine länderübergreifende europäische Studie durchzuführen oder zu unterstützen. Daher wurde die Frage nach Möglichkeiten für so eine Studie am meisten diskutiert. In diesen Diskussionen lag ein besonderes Interesse der ExpertInnen auf **Standards für Programme** und auf **Ergebnismessungen**, die in Evaluierungsstudien verwendet werden. Einen Ausgangspunkt, um die Mindeststandards für Täterarbeitsprogramme und deren Ergebnismessungen zu definieren gab es in der *Convention on Preventing and Combating Violence Against Women and Domestic Violence* („Istanbul Convention“, Europarat, 2011), in einer Veröffentlichung zu Ergebniskriterien von Westmarland, Kelly and Chalder-Mills (2010), oder in den *Guidelines to Develop Standards for Programmes Working With Male Perpetrators of Domestic Violence* (Work With Perpetrators – European Network [WWP], 2008). Es soll auch betont werden, dass sich die Standards nach Meinung der ExpertInnen auf Politiken zur Geschlechtergleichstellung Geschlechtergerechtigkeit stützen sollen.

Auch in Bezug zu Forschung und Evaluierungsstudien ging es um Mindestanforderungen. **Mindeststandards** sind in **jeder länderspezifischen Studie** oder auch in jedem einzelnen Teil einer solchen Studie nötig, um diese Teile zusammenführen zu können. Einige Elemente wurden genannt, die schon bei verschiedenen Gelegenheiten verwendet wurden, z.B. das ökologische Model; das Stufenmodell der Beobachtung (prä, post, follow-up); bestimmte Verfahren/Instrumente (wie z.B. das *Violence and Abuse Inventory*). ForscherInnen haben bereits begonnen, verschiedene relevante Themen zu beschreiben und zu analysieren, z.B. europäischer und nationaler Kontext (gesetzliche Lage, Art der Zuweisung usw.), Tätertypen, Häufigkeit und Art der Gewalt, gewünschte Veränderungsbereiche (das gewalttätige Verhalten des Täters; Kommunikation; Einstellungen und Identität; familiäre Beziehungen usw.). Im Rahmen der Diskussion stellten einige ForscherInnen erste Forschungsfragen (ohne an diesem Punkt der Diskussion Antworten zu liefern), um auf die folgenden Bereiche einzugehen:

- Was trägt am meisten dazu bei, Gewalt zu beenden und die Sicherheit der Opfer zu erhöhen?
- Ändern sich die teilnehmenden Männer? Warum? Wie?
- Was funktioniert für wen?
- Welche (anderen) Faktoren beeinflussen das Leben der Täter (soziales Milieu, Religion, Migrationshintergrund usw.)?

Schließlich sprachen die ForscherInnen die Voraussetzungen an, die dabei helfen können, die Qualität **zukünftiger einrichtungs- und länderübergreifender Evaluierungsstudien** zu verbessern, z.B.: Triangulation der Forschungsmethoden (qualitativ, quantitativ, Beobachtungen); Triangulation von Daten und Theorien; die Berücksichtigung der Perspektiven der Täter, Opfer und Fachleute. Die ExpertInnen erachteten außerdem die Beteiligung von Dachorganisationen und lokalen spezialisierten NGOs (MigrantInnengruppen, LGBTQ Gruppen, religiöse Gruppen) am Forschungsprozess als in höchstem Maße hilfreich.

### 3.3. Vorteile

In allen Interviews und Diskussionen betonten die ExpertInnen die Vorteile von einrichtungsübergreifenden Evaluierungsstudien. Es sollte eine **gute und brauchbare** Vorgangsweise angewandt werden, Gewalt und andere **Ergebnisse** in einer derartigen Studie zu **messen**. Damit wäre so eine Studie nützlich für Fachleute aus allen europäischen Ländern und würde dazu beitragen, die Lücke an Wissen und Forschung über Gewalt im Allgemeinen zu schließen: die derzeitig bestehende Methodologie ist laut einiger ExpertInnen unzureichend. Besonders die Fachleute gaben ihrer Hoffnung Ausdruck, dass eine Evaluierung auch den **Fortbestand bestehender Programme** fördern würde, da manche EntscheidungsträgerInnen eine weitere Finanzierung von positiven Ergebnissen der Programmevaluierung abhängig machen.

Darüber hinaus wiesen die ExpertInnen darauf hin, dass der Einbezug der Perspektive von Opfern und Opferbetreuungseinrichtungen zu **umfassenderen Evaluierungsstudien** führen

werde und das **erweiterte System**, das an der Intervention beteiligt ist, miteinbeziehen könne. Länderübergreifende europäische Evaluierungsstudien könnten auch dazu beitragen, einzelne nationale Täterarbeitsprogramme zu **bewerben**, die den diversen Interessensgruppen noch wenig bekannt sind. Die Ergebnisse der Programme könnten den Fachleuten, die die Qualität ihrer Arbeit verbessern und Fehler vermeiden wollen usw., als **Wissensgrundlage** dienen. Zusätzlich wird die Aufmerksamkeit, die internationalen Studien zu Täterarbeit entgegengebracht wird, auch die Aufmerksamkeit auf Täterarbeitsprogramme und ihre Rolle in manchen Ländern erhöhen. Laut einiger ExpertInnen wird dies dazu beitragen, die Bewerbung auf nationaler und internationaler Ebene effizienter zu machen. Eine länderübergreifende, europäische Evaluierungsstudie, die als europäisches Projekt durchgeführt wird, könnte die einzige Gelegenheit sein, Programme in jenen Ländern zu evaluieren, die bis jetzt aufgrund von fehlenden Mitteln und fehlender Methodologie noch keine Evaluierungen durchgeführt haben.

### 3.4. Herausforderungen

Die ExpertInnen waren sich einig, dass der **Rekrutierungsprozess** von Programmen eine der größten Herausforderungen von länderübergreifenden Evaluierungsstudien darstellt, vor allem dann, wenn keine zusätzlichen finanziellen Mittel zur Verfügung gestellt werden. Die ExpertInnen betonten, dass so eine Studie nur dann erfolgreich sein kann, wenn die Fachleute und ProgrammmitarbeiterInnen von den ForscherInnen davon überzeugt werden können, **Zeit** und **Ressourcen** zu opfern, um sich an der Forschungsarbeit zu beteiligen. Nach Meinung der ExpertInnen könnte das Argument, dass damit Wissenslücken im Bereich Evaluierungsstudien zu Täterarbeit geschlossen werden können, nicht ausreichen, um maßgebliche AkteurInnen zur Mitarbeit zu motivieren. Daher sollten auch andere **Vorteile** einer Mitarbeit an einer Evaluierungsforschung betont werden, wie z.B. externes Training für ProgrammmitarbeiterInnen (nicht nur in Bezug auf Evaluierungsmethoden), internationale Austauschtreffen, Zertifizierungen (durchgeführt z.B. von *Work With Perpetrators - European Network* oder anderen nationalen Organisationen, wie Respect in Großbritannien oder BAG Täterarbeit HG in Deutschland). Nach Meinung der ForscherInnen unter den ExpertInnen wird einer der wichtigsten Vorteile einer Evaluierungsstudie das Feedback von den Klienten sein, das zu einer Verbesserung der Programmqualität und Programmeffektivität führen kann. Außerdem haben Programme, die an einer Studie teilnehmen, höhere Chancen, in Zukunft Fördermittel zu erhalten, wenn ein Evaluierungsprozess abgeschlossen ist und bestätigt, dass das Programm Wirkung zeigt. Die europäische Dimension einer extern geführten Studie würde die Glaubwürdigkeit der Ergebnisse auch auf der nationalen Ebene erhöhen.

Eine weitere Herausforderung ist es, sich in Bezug auf die Verwendung derselben **Evaluierungsmodelle** zu **einigen** (Kriterien, Designs, Methoden, Verfahren/Instrumente). Zum jetzigen Zeitpunkt sind die Programme und praktischen Methoden und auch die Kriterien und Verfahren, die bei der Ergebnismessung verwendet werden, in vielen Ländern unterschiedlich. Demnach besteht die Herausforderung nicht nur darin, dieselben Instrumente auszuwählen und zu verwenden, sondern auch darin, jene zu finden, am allgemeingültigsten und übertragbarsten sind.

### 3.5. Hindernisse

Den ExpertInnen zufolge, gibt es viele Hindernisse bei der Durchführung von länderübergreifender europäischer Forschung zu Ergebnissen von Täterarbeit. Das schwerwiegendste Problem betrifft die **Finanzierung**, da bei der Durchführung einer so komplexen Studie viele Kosten anfallen. Daher ist es für ein derartiges Projekt äußerst wichtig und auch besonders schwierig, finanzielle Unterstützung zu finden. Finanzierung ist nötig, um die Arbeit externer ForscherInnen und andere Projektausgaben (Reisekosten, Ausstattung usw.) zu bezahlen, aber auch, um die zusätzliche Arbeit der ProgrammmitarbeiterInnen zu finanzieren, vor allem dort, wo Fachleute unterbezahlt sind oder sogar ehrenamtlich arbeiten (was in ehemals sozialistischen Ländern häufig der Fall ist) oder auch dort, wo Programmevaluierung nicht Teil der regionalen oder nationalen Förderung ist.

Der **unterschiedliche Kontext** der verschiedenen europäischen Länder ist ein weiteres Hindernis, das sich sowohl auf die Ausgestaltung der Programme als auch auf den zukünftigen Forschungsprozess auswirkt. Unterschiedliche gesetzliche Gegebenheiten können die Auswahl der spezifischen Methoden und die grundlegenden Annahmen, auf denen ein Forschungsprojekt basiert, beeinflussen. In manchen Bereichen können gesetzliche Bestimmungen, die sich auf das Interventionssystem auswirken, Regelungen zum Datenschutz usw. ForscherInnen bei der **Erfassung gewisser Daten** behindern. Ein typisches Beispiel für eine derartige Situation wurde von einer Expertin aus Spanien genannt, wo MitarbeiterInnen von Täterarbeitsprogrammen des Justizsystems keinen Zugang zu den Opfern haben und es somit schwierig oder sogar unmöglich ist, die Perspektive der Partnerinnen miteinzubeziehen.

Ein weiteres Problem ergibt sich aus dem kulturellen Kontext der verschiedenen europäischen Länder, der sich ebenfalls auf den Forschungsprozess auswirken kann; das **Verständnis grundlegender Begriffe** wie z.B. *häusliche Gewalt, Opferrechte, unangemessenes/deviantes Verhalten, Geschlechtergerechtigkeit* usw. können sich unterscheiden und somit eine Herausforderung für die Gestaltung länderübergreifender Forschungsinstrumente darstellen. Außerdem unterscheiden sich die Länder in Bezug auf ihre **wirtschaftliche** Situation, die Rolle der **Religion(en)** (das Ausmaß, in dem die Politik von religiösen Vorstellungen beeinflusst ist, also mehr oder weniger säkulare Länder), die Vielfalt kultureller Hintergründe in Verbindung mit **Migration** usw., aber auch in Bezug auf den Grad der **Implementierung von Geschlechterpolitik** und Gender Mainstreaming, den Grad der Implementierung und Entwicklung von **Täterarbeitsprogrammen** und ihre **Akzeptanz der Wichtigkeit** eines derartigen Interventionssystems.

Bei den Programmen erkennen die ExpertInnen wichtige Unterschiede unter den Ländern in Bezug auf **Wirkmodelle, Erfolgskriterien, Typologien von Männern**, die an Täterarbeitsprogrammen teilnehmen und unterschiedliche Zugänge zum **Geschlecht** der KlientInnen in den Gruppen: Es gibt Programme, die nur mit Männern arbeiten, aber auch solche, die einen koedukativen Zugang haben, d.h. entweder mit mit verschiedenen Gruppen/Zugängen für Täter und Täterinnen arbeiten oder Unterscheidung vorgehen. Zusammenfassend kann man sagen, dass die ExpertInnen der Meinung sind, dass die schwerwiegendsten (methodologischen) Hindernisse für länderübergreifende, europäische

Evaluierungsforschung darin bestehen, dass **Täterarbeitsprogramme** aus vielfältigen Gründen ganz einfach sehr **unterschiedlich** sind.

Schlussendlich haben einige ExpertInnen daran gezweifelt, dass es möglich ist, eine komplexe europäische Studie durchzuführen, weil es immer noch nicht genug Täterarbeitsprogramme gibt (in manchen Ländern, wie z.B. Ungarn, gibt es noch keine Programme; Kutrovátz, 2013), was ein schwerwiegendes Rekrutierungsproblem zur Folge haben könnte.

### 3.6. Was sich aus den ExpertInnen-Fokusgruppen und Diskussionen ergibt

Der Beitrag der ExpertInnen zu den Gruppendiskussionen und ihre Antworten auf verschiedene Fragen wurden zu den folgenden ExpertInnenempfehlungen zusammengefasst:

- Zuallererst muss es eine **Definition** von „Täterarbeitsprogramm“ geben, die im Rahmen von länderübergreifenden Evaluierungsstudien berücksichtigt werden soll. Die üblichen Programme (basierend auf kognitiver Verhaltenstherapie oder einem psycho-edukativen Zugang mit einem Fokus auf Gender) scheinen zu diesem Zeitpunkt die geeignetsten Programme darzustellen. Dies hat auch methodologische Gründe: wir benötigen *ähnliche* Programme in *unterschiedlichen* Kontexten – und nicht *unterschiedliche* Programme in *unterschiedlichen* Kontexten. Es muss bei den Programmen auch möglich sein, die passenden Ergebnisdaten zu erheben, was von der Kooperation bzw. einem koordinierten Zugang abhängt. Außerdem sollte sich die Studie auf Programme konzentrieren, die mit bestehenden Standards und Richtlinien **kompatibel** sind (z.B. die Istanbul Convention).
- **Forschungsprojekt:** In einem neuen, geförderten Projekt zusammen zu arbeiten wurde als der realistischste Weg gesehen, um ein einrichtungs- und länderübergreifendes Evaluierungsprojekt durchzuführen. Das ursprüngliche Konzept, das Instrumentarium „Toolkit“ aus WS 4 für die Evaluierung zu verwenden, ebenso wie die zentrale Erfassung und Analyse von Daten, wurde – zumindest zu diesem Zeitpunkt im Prozess – hinterfragt. Dennoch sollte das Instrumentarium so breit wie möglich eingeführt und die Forschung soweit als möglich auf diesem Instrumentarium aufgebaut werden. Dadurch kann es später möglich sein, zum ursprünglichen Konzept zurückzukehren. Die ExpertInnen betonten den Unterschied zwischen dem Instrumentarium (als Produkt des Workstream 4, zur Ergebnismessung und Ergebniskontrolle) und den Verfahren, die Teil einer Evaluierungsmethodologie sind. Es erscheint schwierig, Evaluierungsprojekte einfach in die tägliche Arbeit einzubinden.
- Zukünftige Studien sollten **umfassend** gestaltet werden und das **System**, das an der Intervention beteiligt ist, miteinbeziehen. Die Perspektive auf die gesellschaftlichen Auswirkungen von Täterarbeitsprogrammen könnten das ursprüngliche Evaluationsdesign zu den Auswirkungen eines Programmes erweitern. Unterschiedliche **Kontexte** in den unterschiedlichen Ländern sollten von Anfang an beachtet werden. Dies umfasst gesetzliche Bestimmungen, Aspekte des Interventionssystems und kulturelle Themen.

- Evaluationskriterien, Designs, Methoden, Verfahren/Instrumente usw. sollten mit bestehenden **Good-Practice-Evaluationsverfahren** verbunden werden und die Ansichten der **ExpertInnen** berücksichtigen.
- Die vielen **Fragen**, die durch Evaluierungsstudien beantwortet werden sollen, sollen systematisiert werden. Vorzugsweise sollte sich diese **Systematisierung** auch auf die verschiedenen **Interessensgruppen** beziehen.
- Evaluierungsmodelle sollten einen hohen Grad an **Übertragbarkeit** aufweisen, um auf längere Sicht viele Projekte in vielen Ländern miteinbeziehen zu können.
- Zur Art des Programms: **Zuweisungswege** (Selbstmelder, von Institutionen zugewiesen, gemischt) könnten als Variable verwendet werden – und nicht als Auswahlkriterium.
- In jedem Fall sollen die **Vorteile** an einer Teilnahme an Evaluierungsstudien während des Rekrutierungsprozesses hervorgehoben werden.

## 4. Evaluierung von Täterarbeitsprogrammen bei häuslicher Gewalt: Wie es weitergehen kann

### 4.1. Ausgangspunkt: Der aktuelle Stand

Wie sich bei der Lektüre der vorhandenen Quellen zur Evaluierung von Täterarbeitsprogrammen bei häuslicher Gewalt und die Perspektiven der ExpertInnen gezeigt hat, musste das ursprüngliche Konzept des Workstream 3 modifiziert werden. Es hat sich herausgestellt, dass es nicht durchführbar ist, große Evaluierungsstudien zu planen, gestalten und durchzuführen, die die zahlreichen Fragen der verschiedenen Interessensgruppen umfangreich beantworten können – zumindest momentan nicht. Stattdessen haben wir ein Konzept für koordinierte Evaluierungen entwickelt, das aus einer Reihe von kleineren Evaluierungsprojekten mit ähnlicher Methodologie besteht und die Koordination dieser Projekte ermöglicht.

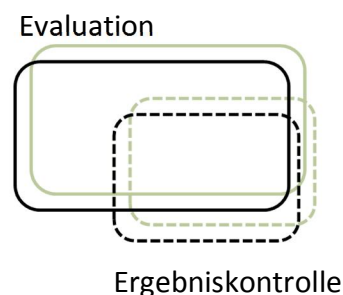
Die Voraussetzungen für länder- und einrichtungsübergreifende Projekte wurden folgendermaßen zusammengefasst:

Die Hauptbeweggründe für so ein Projekt sind es, die große Variabilität nationaler Bedingungen und „Systeme“ in ganz Europa (Praktiken der Institutionen, Opferbetreuung, Täterarbeitsprogramme bei häuslicher Gewalt usw.) zu nutzen. Diese Variabilität kann eine Ressource für Evaluierungsstudien sein, um Täterarbeitsprogramme in verschiedenen Kontexten zu untersuchen. Außerdem wurden viele Programme noch nicht evaluiert, und daher können länder- und einrichtungsübergreifende Evaluierungsstudien eine Gelegenheit sein, um die empirischen Grundlagen von Täterarbeitsprogrammen zu erweitern. Wie oben beschrieben, gibt es bei Täterarbeitsprogrammen eine große Vielfalt unter den Ländern, aber auch ähnliche Programme (z.B. Programme, die auf kognitiver Verhaltenstherapie oder auf dem Duluth-Ansatz basieren). Auch Elemente aus koordinierten Interventionen sind weitverbreitet. In vielen Ländern werden Verfahren zur Ergebnismessung schon verwendet, allerdings gibt es große Unterschiede in den Methodologien, was es wiederum schwierig macht, die bestehenden Studien zu vergleichen. Es gibt Unterschiede und Schwierigkeiten in

Bezug auf Terminologie und Konzepte (z.B. „Abschluss“, „Erfolg“, usw.) und auch in Bezug auf die Designs (z.B. die Zeitpunkte der Messung und die Informationsquellen betreffend); es ist schwierig, Kontrollgruppen einzurichten, und es gibt ethische Bedenken in Bezug auf die Kontrollgruppen; Stichprobenreduktion durch Ausfälle ist in der Praxis ein Hauptproblem, welches sowohl die Männer in den Programmen, als auch die Opfer betrifft. Allerdings kommen Hester et al. (2014) bei ihrem Überblick von Studien zum Schluss, „... it would be possible in the main to take elements from different approaches in order to start developing a robust evaluation methodology.“ (Hester et al., 2014, S. 39)

ExpertInnen erachten es als sinnvoll, das Konzept des länderübergreifenden Zugangs weiter zu verfolgen; die verschiedenen Interessensgruppen erhoffen sich Antworten auf Fragen zur Effektivität von verschiedenen Programmen zur Täterarbeit bei häuslicher Gewalt. Wie verschiedene Interessensgruppen darlegten, sollte der Fokus auf Programmen liegen, die mit bestehenden Standards und Richtlinien vereinbar sind (z.B. Istanbul Convention). Es stellte sich jedoch heraus, dass Evaluierungsstudien, die relevante Fragen beantworten können, großen Aufwand bedeuten und Ressourcen brauchen. Daher bezweifeln ProjektleiterInnen und Fachkräfte, dass sie diese Arbeit in ihre täglichen Routinen einbinden können. Anders ausgedrückt scheint es nicht machbar, sich ohne zusätzliche Mittel in Richtung von internationalen, länder- und einrichtungsübergreifenden Evaluierungsstudien zu bewegen. In der Praxis mag es möglich sein, Interessensgruppen davon zu überzeugen, die Verfahren, die sie für die Bewertung von Ergebnissen bzw. Ergebniskontrolle verwenden, gegen alternative Verfahren auszutauschen. Es bleiben jedoch einige Probleme bestehen, die sich nicht innerhalb der täglichen Praxis der Programme lösen lassen. Es ist wichtig anzuerkennen, dass vernünftige Evaluierungsstudien ein breiteres Konzept haben und dadurch mehr Mittel benötigen als eine reine Ergebniskontrolle. Arbeit und Aufwand von Evaluationsstudien gehen weit über die tägliche Routine von Ergebnismessung hinaus, obwohl Ergebnismessungen einen guten Ausgangspunkt und eine gute Basis darstellen, um Täterarbeitsprogramme bei häuslicher Gewalt für die Forschung vorzubereiten (vgl. Gondolf, 2012; vgl. Abbildung 2). Folglich haben Fachleute betont, dass zusätzliche finanzielle Mittel eine Voraussetzung sind, um an Evaluierungsstudien teilzunehmen.

**Abbildung 2. Sich überschneidende und spezifische Aspekte von (a) Evaluierungsstudien und (b) Verfahren der Ergebniskontrolle.**



Alles in allem kann man zum Schluss kommen, dass es eine Grundlage für länderübergreifende Evaluierungsprojekte gibt. Allerdings ist eine Harmonisierung der Methodologie absolut vonnöten. Täterarbeitsprogramme bei häuslicher Gewalt und

Opferbetreuungseinrichtungen müssen Teil von Evaluierungsprojekten sein. Um die Erfolgchancen zu erhöhen, sollen zukünftige Evaluierungsstudien daher als separate Projekte geplant werden, die ausreichend gefördert sind, um die zusätzlichen Kosten abzudecken, die bei den beteiligten Organisationen anfallen.

#### 4.2. In welche Richtung wir gehen sollen

Idealerweise wird die fragmentierte Situation bei den Evaluierungen von Täterarbeitsprogrammen bei häuslicher Gewalt in einigen Jahren überwunden sein, indem die Evaluierungsmethodologie harmonisiert wurde, damit Ergebnisse verschiedener Studien miteinander verglichen werden und einander ergänzen können. Wir stellen uns vor, dass es eine koordinierende Einrichtung gibt, die zukünftige Evaluierungsprojekte von Täterarbeitsprogrammen im Bereich Design, Methoden und Verfahren sowohl verbreitet, als auch unterstützt und beaufsichtigt. Diese Einrichtung soll auch einen Überblick darüber haben, welche Fragen gestellt und beantwortet werden sollen. Dies wird kein vollständiges *Forschungsprogramm zur Evaluierung von Täterarbeitsprogrammen bei häuslicher Gewalt* sein, da eine Reihe von mehrjährigen Studien nicht in umfassender Weise vorhergesehen und geplant werden können – vor allem aus Gründen der eingeschränkten Finanzierung. Es ist jedoch wahrscheinlich, dass es diverse Evaluierungsstudien geben wird – manche auf nationaler Ebene, andere auf internationaler Ebene (z.B. als EU-Projekte). Es muss möglich sein, diese Studien zu vergleichen und miteinander zu verknüpfen.

Aus verschiedensten Gründen schlagen wir vor, mit dem *Work With Perpetrators – European Network* (WWP-EN) als koordinierende Institution und dem *WAVE-Network als Unterstützung zu arbeiten*: WWP-EN hat die führende Rolle bei der Koordinierung von Täterarbeitsprogrammen in der EU und darüber hinaus übernommen. Als Dachorganisation von Täterarbeitsprogrammen, gemeinsam mit kooperierenden Opferbetreuungseinrichtungen, wurden bereits Schritte in Richtung eines Standards für die konkrete Täterarbeit eingeleitet. Es ist wahrscheinlich, dass dies auch in Zukunft passieren wird. Es ist einfach ein logischer Schritt, diese Arbeit zu evaluieren und auch **Standards** für diese **Evaluierungsarbeit** zu entwerfen (vgl. unten).

Dies stimmt mit dem überein, was Gondolf „practitioner-led evaluations“ (2012, S. 81) nennt bzw. mit Evaluierungsprojekten, bei denen die Initiative aus dem praktischen Bereich ausging oder davon mitgestaltet oder beeinflusst wurde. Gondolf streicht hervor, dass die Zusammenarbeit von ForscherInnen und Fachleuten in vielen Fällen schwierig war, weil jede Seite die andere nur zum Teil verstand. Hier gibt es ein großes Verbesserungspotential.

#### 4.3. Standards für Evaluierungsprojekte

Wie wir aus den Workstreams 1 und 2 und der Lektüre der Fachliteratur gelernt haben, sind methodologische Standards und eine einheitliche Terminologie essentiell für eine sinnvolle Entwicklung zukünftiger Evaluierungsprojekte. Täterarbeit im Allgemeinen und Täterarbeitsprogramme bei häuslicher Gewalt im Speziellen sind sehr spezifische Aufgaben, weswegen Verfahren aus anderen Bereichen (z.B. Evaluierung von Psychotherapie) nicht



einfach übernommen werden können. Einige der spezifischen Aspekte der Evaluierung von Täterarbeitsprogrammen bei häuslicher Gewalt sind:

- *Konzepte und Terminologie*: Es gab bis jetzt umfangreiche Diskussionen darüber, worauf Täterarbeitsprogramme bei häuslicher Gewalt abzielen sollen. Diese Diskussionen zeigten, dass unterschiedliche Begriffe und Konzepte (wie z.B. „Erfolg“) verwendet werden. Verschiedene Interessensgruppen haben ihre Positionen und Interessen in Bezug auf die Ergebnisse von Täterarbeitsprogrammen bei häuslicher Gewalt benannt. Diese Positionen müssen bei der Planung von Evaluierungsstudien berücksichtigt werden.
- *Informationsquellen*: ExpertInnen im Feld der Täterarbeit erachten die Information, die von den Männern selbst kommt, als unzureichend (vgl. z.B. Gondolf, 2012). Information von Partnerinnen, Polizei oder Gerichtsakten sind notwendige Ergänzungen zur benötigten Information. Die Information der Partnerinnen wird als zuverlässigste und sinnvollste Information erachtet und sollte somit zentraler Indikator sein. Dies wirft allerdings Fragen zur Ethik auf, die berücksichtigt werden müssen (vgl. Downes, Kelly & Westmarland, 2014).
- *Beobachtungszeitpunkte*: Um sich einerseits mit den relativ hohen Ausfallraten (sowohl die Täter als auch die Opfer betreffend) zu beschäftigen und andererseits die Nachhaltigkeit der durch das Programm erreichten Veränderungen bewerten zu können, reicht ein einfaches Prä-Post-Design nicht aus, um Täterarbeitsprogramme bei häuslicher Gewalt zu evaluieren. Hester et al. (2014) haben in diesem Kontext ein Schema von Beobachtungszeitpunkten vorgeschlagen. Sie haben vier Zeitpunkte definiert; T0: wenn ein Mann einem Programm zugewiesen wird; T1: wenn das Programm beginnt; T2: während des Programms; T3: am Ende des Programms; T4: Follow-up, beispielsweise 6 Monate nach Abschluss des Programms. Dieses Schema kann gemeinsam mit Vorgaben zur Information, die zu jedem dieser Zeitpunkte erhoben werden soll (wer muss gefragt werden, welche Verfahren sollen verwendet werden), einen zentralen Standard für die Evaluierung von Täterarbeitsprogrammen darstellen. Gemeinsam mit einer ausreichenden Beschreibung der Männer zum Zeitpunkt T0 (z.B. sozio-demographische oder psychologische Merkmale) und dem Zuweisungsweg kann das Thema *Ausfallraten* viel besser untersucht werden als dies heute möglich ist.

#### 4.4. Verbreitung der Standards

Nach der Entwicklung derartiger Standards ist es notwendig, sie zu verbreiten und in Evaluierungsstudien zu verwenden. Im Rahmen des Workstream 3 wurden verschiedene Arten zur Verbreitung solcher Standards diskutiert. Die folgenden Möglichkeiten wurden genannt:

- Eine Taskforce innerhalb des *WWP-EN* Netzwerks, die für die Themen der Evaluierung verantwortlich ist, könnte die Schlüsselrolle in der Harmonisierung von Evaluierungsmethodologie spielen, indem sie bewährte Methoden der Evaluierungsforschung von Täterarbeitsprogrammen sammelt und die einzelnen Evaluierungsstudien koordiniert und steuert. Zukünftige Evaluierungsprojekte werden weiterhin mit den Programmen in der Praxis kooperieren müssen. Daher ist

es wahrscheinlich, dass Evaluierungsprojekte auf der Zusammenarbeit mit Mitgliedsorganisationen des WWP-EN beruhen, die von der WWP-EN-Taskforce für Evaluierungsstandards sensibilisiert wurden.

- In gleicher Weise werden internationale Netzwerke von Opferbetreuungseinrichtungen, vor allem das WAVE Netzwerk, eine Schlüsselrolle spielen. Evaluierungen von Täterarbeitsprogrammen bei häuslicher Gewalt können nicht auf die Perspektive der Opfer verzichten. Daher werden derartige Projekte mit WWP-EN Mitgliedsorganisationen und auch mit WAVE Mitgliedsorganisationen in Kontakt stehen und zusammenarbeiten. Da WAVE und WWP-EN zusammenarbeiten, können diese beiden europäischen Dachorganisationen durch ihre Mitgliedsorganisationen einen maßgeblichen Einfluss auf die Harmonisierung der Evaluierungsmethodologie nehmen. Außerdem gibt es die Möglichkeit, (nationale und internationale) Evaluierungsstudien innerhalb des WWP-EN zu initiieren. Vor allem bei länderübergreifenden EU-Projekten ist das WWP-EN die ideale europäische Organisation, um Projektanträge im Bereich der Evaluierungsforschung zu koordinieren.
- Verbreitung der Methodologiestandards durch ein von der EU finanziertes Evaluierungsprojekt ist der nächste Meilenstein. Innerhalb dieses zukünftigen Projekts kann die konkrete Methodologie auf der Basis der Machbarkeitsbewertung des Workstream 3 entwickelt werden. Dies ist durch eine Sammlung geeigneter Programme in verschiedenen Ländern möglich. Diese Programme werden aus den Daten des Workstream 1 und des WWP-EN ermittelt. Sobald ein Projekt den ersten Schritt in Richtung einer länderübergreifenden Evaluierung von Täterarbeitsprogrammen gesetzt hat, können andere folgen und auf den Erfahrungen des Pilotprojekts aufbauen.
- Förderung von forschungsähnlichen Praktiken durch das Instrumentarium aus dem Workstream 4. Obwohl es bei diesem Instrumentarium um die Ergebniskontrolle in der täglichen Praxis der Täterarbeitsprogramme geht, trägt es auch zur Harmonisierung der Datenerfassung bei, worauf zukünftige Studien verschiedener Länder aufbauen können. Wenn mehrere Programme das Instrumentarium benutzen, werden sie forschungsähnliche Praktiken in ihre tägliche Arbeit aufnehmen, die den Weg für ein Evaluierungsprojekt ebnen. Die Ergebniskontrolle wird so die Programmteams auf die Forschung vorbereiten.
- EU-Programme zur EU-weiten Harmonisierung von Forschung und Praxis ergeben auch Möglichkeiten zur Harmonisierung von Methoden im Bereich der Evaluierungsforschung. Es gibt EU-Programme, die an der Entwicklung einer Europäischen Forschungslandschaft (*European Research Area*) arbeiten, z.B. COST. Auf der praktischen Ebene werden die Netzwerke WWP-EN und WAVE eine führende Rolle bei der Verbreitung methodologischer Standards bei der Evaluierung spielen, indem sie dabei helfen, Programme auszuwählen und sie ihre Mitglieder dazu ermutigen, an länderübergreifenden Evaluierungsstudien teilzunehmen.

#### **4.5. Ein umfassender Zugang zur Evaluierung von Täterarbeitsprogrammen bei häuslicher Gewalt und die Rolle länderübergreifender Studien**

Verschiedene Arten von Evaluierungsstudien können verschiedene Fragen beantworten. Wichtig dabei ist, dass diese Studien verglichen und miteinander verknüpft werden können. Dies kann sichergestellt werden, indem dieselbe Methodologie verwendet wird (oder sie zumindest im Kern vergleichbare Verfahren und Instrumente haben). Wir schlagen vor, die verschiedenen Arten von Evaluierungsstudien näher zu betrachten, und zwar aus der Perspektive der Kategorien, die unten umrissen werden. (Weitere detaillierte Informationen darüber und über andere Aspekte werden im Rahmen des IMPACT Projektes in separaten Beiträgen veröffentlicht werden).

##### **4.5.1. Grundlegende Studien in einem einzigen Kontext**

Es handelt sich hierbei um Studien, die die Auswirkungen eines Täterarbeitsprogrammes bei häuslicher Gewalt als Teil einer koordinierten Intervention evaluieren. Derartige Studien werden nicht unter Laborbedingungen, sondern in der Praxis durchgeführt. Daher wenden sie eher quasi-experimentelle Designs an, weil eine Zufallszuordnung von Forschungseinheiten (z.B. Männer) zu Behandlungsbedingungen (Intervention/keine Intervention) oft aus methodologischen und ethischen Gründen nicht möglich ist. Die Ausfälle unter den Teilnehmern oder andere Gruppen von Männern, die die Intervention nicht abgeschlossen haben, werden als Kontrollgruppen verwendet (z.B. Männer, die nicht in die Programme zugelassen oder später ausgeschlossen wurden). Diese Gruppen wurden mit jenen verglichen, die die Programme abgeschlossen haben.

Wir möchten in diesem Bereich drei Elemente als Teil einer zukünftigen Evaluierungsmethodologie vorschlagen, die noch ausgearbeitet werden müssen z.B. durch die *WWP-EN Evaluation Taskforce*: (a) Beobachtungen/Zeitpunkte; (b) Erfolgsindikatoren; (c) Design.

##### **(a) Beobachtungen / Zeitpunkte**

Hester et al. (2014) haben in ihrer Analyse europäischer Evaluierungsstudien ein Schema vorgelegt, das Folgendes beschreibt: „... shape and size of the sample at the different stages of the evaluation process“ (S. 17). Dieses Schema sollte auch in Evaluierungsstudien von Täterarbeitsprogrammen bei häuslicher Gewalt verwendet werden, um die verschiedenen Zeitpunkte der Beobachtung zu definieren. Dieses Schema besteht aus fünf Zeitpunkten und Vorschlägen für Informationsquellen zum jeweiligen Zeitpunkt:

- T0: „Vor dem Beginn der Intervention“. Es ist wichtig zu wissen, welche Männer den Programmen zugewiesen werden (sozio-demographische Daten, psychologische Merkmale, Substanzmissbrauch, Art der Zuweisung usw.), und ob sie das Programm beginnen oder nicht. Der Vergleich von Stichproben zu den Zeitpunkten T0 und T1, zusammen mit den Gründen für einen eventuellen Ausfall, können sinnvolle Informationen für die Entwicklung von Interventionen für Männer, die das Programm nicht beginnen, bringen. Viele dieser Informationen können von den Opfern, manche von den Männern, andere von den Fachleuten kommen. (Für Information bezüglich

ethischer Aspekte von Evaluierungsprogrammen vgl. Downes, Kelly & Westmarland, 2014.) Informationen zu den Indikatoren, die von den **Opfern** als Zeichen einer erfolgreichen Intervention erachtet werden (vgl. unten), sollen zu T0 erfasst werden. Diese Indikatoren werden zentrale Ergebnismessungen zu den Zeitpunkten T3 und T4 darstellen.

- T1: „Am Beginn der Intervention“. Der Vergleich der Stichproben zu T0 und T1 wird Auskunft über die erste Welle der Ausfälle geben. Der Vergleich zu T1 und T2 hingegen wird zeigen, welche Männer im Programm bleiben und welche es verlassen. Von T0 bis T3 werden Veränderungen der Lebenssituation und Informationen über Kontextvariablen wichtig sein, um die verschiedenen Ausfälle interpretieren zu können. Bei den **Männern** wird man sich zu diesem Zeitpunkt auf Variablen konzentrieren, die zu T1 beobachtet wurden (z.B. Verhalten, Einstellungen, Motivation, Risikofaktoren usw.), und die sich durch das Programm ändern sollen (als Zielgrößen für die Behandlung). Hier können auch wieder die **Opfer** zu den Indikatoren einer erfolgreichen Intervention befragt werden.
- T2: „Während der Intervention“. Es wird empfohlen, Messungen während der Intervention zu wiederholen. Die Entwicklung der Indikatoren über die Zeit kann wichtige Informationen liefern, sowohl die Männer als auch die Opfer betreffend. Außerdem können so prozessbezogene Fragen angesprochen und mit den Ergebnissen verknüpft werden.
- T3: „Bei Abschluss der Intervention“. Der Vergleich der Behandlungsgruppe mit der Kontrollgruppe zu T1 (vor der Behandlung) und T3 (nach der Behandlung) stellt den Kern von Studien zur Evaluierung der Auswirkungen von Interventionen dar. Wieder sollte die Analyse mindestens zwei Perspektiven beinhalten: die Perspektive der Opfer in Bezug auf die Erfolgsindikatoren und die Perspektive der Männer im Programm bzw. der Kontrollgruppe, in Bezug auf die Ergebnismessung für die Behandlung selbst. Information von Fachleuten und zusätzliche Information aus Akten können ebenso zu T3 erfasst werden.
- T4: „Follow-up“. Die Stabilität der Veränderungen kann bei einem Follow-up erhoben werden. Idealerweise findet das Follow-up sechs Monate nachdem die Männer das Programm abgeschlossen haben statt. Wieder sollten die Opfer und die Männer, die in Programmen waren, befragt werden. Fachleute und Akten können als weitere Informationsquellen dienen.

Eine Analyse der Ausfälle kann zu jedem Zeitpunkt durchgeführt werden, sowohl für die Männer als auch für die Opfer, sofern relevante Informationen (z.B. sozio-demographische Variablen, Risiko, Behandlungsbedarfe etc.) zu T0 erhoben wurden.

#### (b) Designs

Randomisierte kontrollierte Studien (RKS) werden als das höchste Niveau (als sogenannter „gold standard“) in klinischer und psychologischer Forschung angesehen. RKS beruhen auf der zufälligen Zuordnung von Forschungseinheiten (z.B. Männer) zu Behandlungsbedingungen (z.B. *Intervention* verglichen mit *keiner Intervention*). Es gibt viele Schwierigkeiten in Verbindung mit diesem Vorgehen (z.B. Probleme der Ethik). Wenn RKS umgesetzt wurden, wurden sie verschiedentlich kritisiert (vgl. Gondolf, 2002, 2012). Sogar im Rahmen des Strafrechtssystems ist es oft schwierig, RKS durchzuführen (vgl. z.B. Dobash

et al., 2000), obwohl man wohl die besten Bedingungen für RKS in Gefängnissen vorfindet (vgl. unten).

### Abbildung 3. Design einer *Kontrollierten Randomisierten Untersuchung (RKS)*

Zufalls- Zuteilung von Individuen zu Gruppen		T1 (prä)	(T2, während IV)	T3 (post)	(T4, Follow-up)
Interventionsgruppe		O	X O	O	O
Kontrollgruppe		O		O	O

Anmerkung. IV... Intervention.

Als Folge wurden in vielen Studien die Männer, die ausfielen, nicht anwesend waren oder aus dem Programm ausgeschlossen wurden, als Kontrollgruppen definiert, während Abschießer von Täterarbeitsprogrammen bei häuslicher Gewalt als Behandlungsgruppen definiert wurden. Folglich sind die Behandlungs- und Kontrollgruppen nicht äquivalent, was eine Gefahr für die interne Validität solcher Studien darstellt (vgl. Shadish, Cook & Campbell, 2002). In seiner berühmten Evaluierungsstudie hat Gondolf (2002) im Vergleich zu den Ausfällen nur kleine Auswirkungen bei den Abschießern festgestellt. Um das Problem der nicht äquivalenten Kontrollgruppe zu lösen, hat Gondolf spezielle statistische Verfahren (*Propensity Score Analysis* und *Instrumental Variable Analysis*) verwendet, um so die Ergebnisse anzupassen. Diese Verfahren führten zu höheren Effektgrößen. Das Prinzip derartiger Anpassungsverfahren funktioniert wie folgt: Der Unterschied zwischen den Resultaten der Behandlungsgruppe und der nicht äquivalenten Kontrollgruppe wird statistisch korrigiert, indem die Unterschiede zwischen den Gruppen (die zu Beginn der Studie feststehen müssen) als eine Art von Kontrollvariablen verwendet werden. Infolgedessen erhält man eine Effektgröße, die das reflektiert, was als Effekt erwartet werden kann, wenn man annehmen würde, dass die beiden Gruppen so ähnlich wie möglich sind. Diese korrigierte Effektgröße wird von ExpertInnen im statistischen Bereich als zutreffender angenommen als die ursprüngliche nicht korrigierte (und die nicht korrigierte Effektgröße wird als eigentlich nicht zutreffend angesehen).

Viele frühere Evaluierungsstudien haben sich auf Männer in den Programmen konzentriert. Wenn Opfer als Informationsquelle über die Wirkungen des Programms miteinbezogen wurden, hatten ForscherInnen oft zuerst Kontakt mit den Tätern und haben danach versucht, Kontakt zu den jeweiligen Opfern aufzunehmen. Dieser Zugang kann auch in Zukunft verfolgt werden. Wie Gondolf (2002) in seiner Studie zeigt, konnten bis zu 60% der ursprünglichen Opfer für das Follow-up kontaktiert werden.

Es gibt jedoch auch alternative Zugänge, so wie den folgenden, der vom MIRABAL-Projekt<sup>4</sup> inspiriert wurde. Der Fokus kann von den Männern in Programmen auf die Opfer in

<sup>4</sup> Die Ergebnisse des MIRABAL-Projektes werden Anfang 2015 präsentiert. Der Zugang und die Methodologie des Projektes sollten von einer zukünftigen WWP-EN Evaluierungstaskforce aufgenommen werden.

<https://www.dur.ac.uk/criva/projectmirabal/>

Opferbetreuungseinrichtungen verlagert werden. Nachdem es geographische Gebiete mit Opferbetreuungseinrichtungen gibt – aber mit und ohne Täterarbeitsprogramme – kann der zusätzliche Nutzen eines Täterarbeitsprogramms untersucht werden, wenn die Ergebnisse der Opfer in dieser Region oder Stadt mit einer anderen Region/Stadt verglichen werden, die *kein* Täterarbeitsprogramm bei häuslicher Gewalt haben (Opferbetreuungseinrichtungen gibt es in beiden Gebieten). In diesem Design werden die Opfer und auch deren Partner über einen Zeitraum (T1-T4) betreut und begleitet. Die Partner werden einem Täterarbeitsprogramm zugewiesen, wenn eines verfügbar ist (Behandlungsgruppe), oder keinem Täterarbeitsprogramm zugewiesen, weil es keines gibt (Kontrollgruppe). Zum Zeitpunkt T0 ist es möglich, auf der Grundlage der Informationen der Opfer über ihre Partner eine Art von Matching der Behandlungs- und Kontrollgruppen durchzuführen. Obwohl es nützlich und sinnvoll erscheint, die Perspektive der Männer in die Evaluierung einzubeziehen und so zu verfahren wie es oben beschrieben ist („Beobachtungen / Zeitpunkte“), könnten wir sogar darüber nachdenken, nur die Opferperspektive und einige zusätzliche Informationen von den Fachleuten über den Ausfall, die „Dosis“ oder den Abschluss des Programms zu berücksichtigen. In diesem Fall wird das Täterarbeitsprogramm als „Blackbox“ angesehen, das einer Gruppe von Tätern in einer gewissen Dosis (je nach Dauer der Teilnahme) verabreicht wird. Die Behandlungsgruppe (Opfer, deren Partner am Programm teilnehmen) wird dann mit der Kontrollgruppe (Opfer, deren Partner nicht am Programm teilnehmen) verglichen. Als Voraussetzung sollten die beiden Regionen große Ähnlichkeiten aufweisen, damit die Unterschiede zwischen Behandlungs- und Kontrollgruppe dem Fehlen oder Vorhandensein eines Täterarbeitsprogramms zugeschrieben werden können. Obwohl die detaillierten Verfahren komplexer<sup>5</sup> sind, ist das grundsätzliche Konzept, ein Täterarbeitsprogramm mittels verschiedener Ergebnisse der Opfer aus verschiedenen Regionen (mit/ohne Täterarbeitsprogramm), die an einem Opferbetreuungsprogramm teilnehmen, vielversprechend.

Auf ähnliche Weise können auch andere Designs entwickelt werden, z.B. ein Zeitreihendesign mit einer Unterbrechung durch ein bestimmtes Ereignis, z.B. werden Opfer über einen längeren Zeitraum, der mit T0-T5 übereinstimmt, betreut; zu einem gewissen Zeitpunkt wird dann in der Region ein Täterarbeitsprogramm bei häuslicher Gewalt eingerichtet. Die Hypothese ist, dass sich die Ergebnisse, die mit der Lebenssituation der Opfer verknüpft sind, systematisch verbessern, sobald ein Partner einem Täterarbeitsprogramm zugewiesen wird. Zeitreihendesigns eignen sich auch für die Einzelfallforschung. Einzelfallmethodologie ist auch auf der statistischen Ebene eine vielversprechende Entwicklung und kann im Kontext der Evaluierung von Täterarbeitsprogrammen verwendet werden. Die Vorstellung ist verlockend, die Auswirkungen von Täterarbeitsprogrammen auf Basis von Einzelfalldaten zu untersuchen. Dieser Zugang ist besonders nützlich, wenn verschiedene Personen unterschiedliche Behandlungen und Behandlungsintensitäten („Dosen“) erhalten, je nach den individuellen Behandlungsbedarfen.

---

<sup>5</sup> Zum Beispiel müssen Gruppen post-hoc definiert werden: Ausfälle auf Seiten der Opfer und auch der Täter; Abschließer und (Ex-)Partnerinnen von Abschließern; genauso bei Ausgeschlossenen und nicht Zugelassenen; etc.

Im MIRABAL Projekt wurde angekündigt, einen weiteren vielversprechenden quantitativen Zugang in die Evaluierung von Täterarbeitsprogrammen einzubringen, und zwar die *Überlebensanalyse*. Diese Art von Analyse kommt ursprünglich aus der medizinischen Forschung und der Berechnung von Lebenserwartungen, wird aber auch in einer Vielzahl anderer Bereiche verwendet. Überlebensanalyse geht von Sterbetabellen aus und berechnet daraus kumulative Wahrscheinlichkeiten für das Überleben von Bevölkerungsgruppen. Diese Wahrscheinlichkeiten, die üblicherweise im Zeitverlauf dargestellt werden, können für unterschiedliche Untergruppen erstellt werden, z.B. Personen, die unterschiedliche Behandlungen bekommen. Außerdem kann die Wirkung von Einflussgrößen festgestellt werden. Wir empfehlen, dass eine zukünftige WWP-EN Evaluierungstaskforce diesen innovativen Zugang übernimmt.

Die Ergebnisse grundlegender Evaluierungsstudien, wie sie hier beschrieben sind, werden Aussagen über zusätzliche Auswirkungen von Täterarbeitsprogrammen bei häuslicher Gewalt in koordinierten Interventionen darstellen (im Gegensatz zu „koordinierten Interventionen ohne ein Täterarbeitsprogramm“).

Unabhängig vom umgesetzten Design, können unterschiedliche Studien nur dann verglichen werden, wenn das Schema T0-T5 konsequent verwendet wird (um zu wissen, welche Männer in die bestimmten Programme zugelassen werden, wer ausfällt und wann), und wenn dieselben Ergebnismessungen verwendet werden.

#### (c) Erfolgskriterien und Ergebnismessungen

In einem vorbereitenden Forschungsprojekt für das MIRABAL Projekt haben Westmarland, Kelly und Chalder-Mills (2010) die folgenden „Erfolgskriterien“ für Täterarbeitsprogramme vorgeschlagen:

- Verbesserte Beziehung zwischen den Männern in den Programmen und den (Ex)-Partnerinnen; Respekt und bessere Kommunikation
- Erweiterter Handlungsspielraum für (Ex)-Partnerinnen, ihre eigene Stimme wiederfinden, die Möglichkeit, Entscheidungen zu treffen, erhöhtes Wohlbefinden
- Sicherheit und Freiheit von Gewalt und Misshandlungen für Frauen und Kinder
- Sichere, positive und geteilte Kindererziehung
- Erhöhtes Bewusstsein der Männer über sich selbst und andere und Verstehen der Auswirkung von Gewalt auf Partnerinnen und Kinder
- Sichere, gesündere Kindheit, Kinder fühlen sich gehört und spüren, dass man sich um sie kümmert

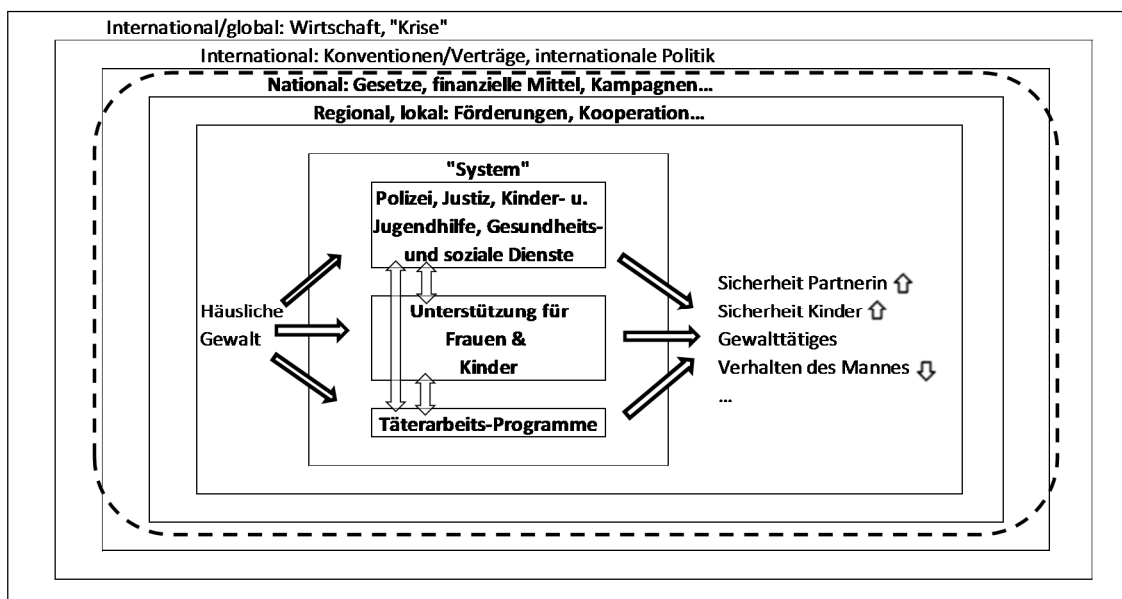
Diese Kriterien wurden dazu genutzt, konkrete Fragen und Themen für die MIRABAL Studie zu formulieren (vgl. Westmarland & Kelly, 2012). Die Perspektiven der Opfer und Kinder stellen bei diesem Zugang, die Auswirkungen eines Täterarbeitsprogrammes bei häuslicher Gewalt zu evaluieren, die wesentlichste Informationsquelle dar. Idealerweise wird diese Information mit der Perspektive der Männer in den Programmen ergänzt, wobei die Ergebnismessungen wie oben dargestellt verwendet werden. Fachleute und Akten können zusätzliche Informationen bringen, die dazu genutzt werden, Fragen zu beantworten, die außerhalb einer Ergebnisevaluierung von Täterarbeitsprogrammen stehen. Um zukünftige

Evaluierungsstudien in einem bewährten Beispiel zu verankern, empfehlen wir, dass eine zukünftige WWP-EN Evaluierungs-Taskforce diese Indikatoren berücksichtigt.

#### 4.5.2. Einrichtungs-und kontextübergreifende Studien

In den *Studien in einem einzigen Kontext* lag der Fokus darauf, die Wirkung von Täterarbeitsprogrammen bei häuslicher Gewalt festzustellen, wenn Bedingungen zweier Gebiete ziemlich ähnlich waren. Nun wenden wir uns einem anderen Fall zu: *Ähnliche* Täterarbeitsprogramme werden in *unterschiedlichen Bedingungen* durchgeführt. Abbildung 4 zeigt, wie man sich diese Konstellation vorstellen kann. Der Kontext eines Täterarbeitsprogramms wird von relevanten, lokalen Komponenten bestimmt, die Bestandteil des Interventionssystems bei häuslicher Gewalt sind (Polizei, Gerichte, Opferschutz usw.), und auch von Bedingungen auf höheren Ebenen, z.B. die Finanzierungssituation auf regionaler Ebene, oder Gesetze, Kampagnen und präventive Maßnahmen auf der nationalen Ebene. Auch internationale und globale Bedingungen können auf die Ergebnisse des Systems Einfluss nehmen.

**Abbildung 4. Ein Analyserahmen für verschiedene Arten von Evaluierungsstudien von Täterarbeitsprogrammen bei häuslicher Gewalt.**



Anmerkung. Fälle von häuslicher Gewalt treffen auf ein Interventionssystem, das aus verschiedenen wechselwirkenden Komponenten besteht. Die Ergebnisse des Interventionssystems sind von jeder Komponente und ihrer Wechselwirkung abhängig, weiters von regionalen, nationalen und internationalen Bedingungen. Aus der Perspektive eines Täterarbeitsprogramms bei häuslicher Gewalt konstituieren die anderen Komponenten des Systems und die Bedingungen auf höherer Ebene (national, international) den *Kontext* des Programms.

In den *Studien in einem einzigen Kontext* sind wir davon ausgegangen, dass der Kontext konstant ist und es in einem Land zwei Gebiete gibt – mit und ohne Täterarbeitsprogramme



– die sich im gleichen oder ähnlichen lokalen und regionalen Kontext befinden. In den einrichtungs- und kontextübergreifenden Studien nehmen wir an, dass sehr ähnliche Täterarbeitsprogramme unter verschiedenen Bedingungen durchgeführt werden, wovon bei einigen Ländern ausgegangen werden kann. Die Ergebnisse des Workstream 1 zeigen, dass es ähnliche Programme in verschiedenen Ländern gibt – vermutlich deshalb, weil sich viele Programme aus denselben Wurzeln heraus entwickelt haben, so wie dem Duluth-Modell oder dem CHANGE Curriculum. Wenn man in zwei oder mehreren Ländern sehr ähnliche Täterarbeitsprogramme findet, die idealerweise dieselben Zuweisungswege aufweisen (vom Gericht zugewiesen, Selbstmelder oder beides), können die Resultate der Interventionen eines gesamten Systems verglichen werden. In diesem Fall können wir annehmen, dass die Funktionsweise des Programms konstant, der Kontext aber anders ist, was zu unterschiedlichen Ergebnissen führt. In solch einem Design werden die *Rolle von Kontext* und seine *Auswirkung auf die Ergebnisse des Programms* verglichen und untersucht. Genau bei dieser Frage ist ein länderübergreifender Zugang sinnvoll.

Um so eine Studie mit anderen Evaluierungen zu verknüpfen, muss in allen Studien dieselbe Kernmethodologie angewandt werden. Als Ergebnis einer einrichtungs- und kontextübergreifenden Studie sind Aussagen über den Einfluss des Kontexts auf die Ergebnisse einer bestimmten Art von Täterarbeitsprogramm zu erwarten.

Wenn der Kontext genauer analysiert werden soll und nicht nur als „Blackbox“ angesehen wird (mit einer bekannten Gesamtwirkung, aber mit unbekannter Wirkung der einzelnen Komponenten), wird die Herausforderung darin liegen, den *Kontext zu operationalisieren*, um den Einfluss der Komponenten auf die Ergebnisse des Täterarbeitsprogramms festzustellen. Von der Abbildung 4 ausgehend ist die Aufgabe, den Kontext in Variablen für eine quantitative Analyse umzuwandeln. Zahlreiche Komponenten müssen „übersetzt“ werden, um sie in quantitative statistische Verfahren integrieren zu können:

- Auf der lokalen Ebene: Unterstützung von betroffenen Frauen und Kindern; die Arbeit von Polizei, Strafrechtssystem, Kinder- und Jugendwohlfahrtssystem, Gesundheits- und Sozialeinrichtungen;
- Auf der regionalen und nationalen Ebene: gesetzliche Maßnahmen, Maßnahmen bezüglich Gewaltprävention, Finanzierungssituation; allgemeine Situation in Bezug auf Geschlechtergerechtigkeit.

Bis jetzt fehlt eine derartige Operationalisierung, und diese wird auch erheblicher Ressourcen bedürfen. Als Vorbereitung für zukünftige *einrichtungs- und kontextübergreifende Studien* wollen wir in diesem Zusammenhang nur einige Hinweise darauf geben, wie eine derartige Operationalisierung durchgeführt werden könnte.

Auf der Ebene des Interventionssystems wurde festgestellt, dass in Fällen der von Gericht zugewiesenen Täter die Reaktionen der Gerichte die Ergebnisse eines Täterarbeitsprogramms beeinflussen (z.B. Gondolf, 2002). Um diesen Einfluss zu operationalisieren, können die Perspektiven von Männern in den Programmen, Fachleuten, Opfern oder ExpertInnen als Ausgangspunkte genommen werden. Wenn wir die Perspektive von *Männern*, die *von Gerichten zugewiesen* wurden und am *Programm teilnehmen*, z.B. durch die *Fachleute* bewerten lassen, könnten die folgenden Indikatoren verwendet werden:

#### A. Erfahrungen und Erwartungen der Klienten bezüglich Gerichtsreaktionen

- „Bei Nichteinhaltung, gibt es eine schnelle und folgenreiche Reaktion des Gerichts (oder eine solche ist zu erwarten).“ (Wert=1)
- „Bei Nichteinhaltung gibt es langsame oder unberechenbare Reaktionen des Gerichts (oder eine solche ist zu erwarten).“ (0,5)<sup>6</sup>
- „Bei Nichteinhaltung gibt es keine Reaktion des Gerichts (oder eine Reaktion ist nicht zu erwarten).“ (0)

#### B. Subjektive Bedeutung angekündigter oder verhängter Konsequenzen

- „Die Entscheidungen des Gerichts sind bedeutend / ernstzunehmend, aus der Sicht des Klienten.“ (1)
- „Entscheidungen des Gerichts haben für den Klienten zum Teil negative Bedeutung.“ (0,5)
- „Entscheidungen des Gerichts sind aus der Sicht des Klienten bedeutungslos.“ (0)

In den einzelnen Fällen ist Information über A, B, A\*B für alle relevanten Komponenten eines Systems und für unterschiedliche Perspektiven (Männer, Fachleute, Opfer) nötig. Auf diese Weise kann pro Fall viel Information über den Kontext in eine Datenmatrix eingetragen werden, die verwendet wird, um z.B. eine Mehrebenen-Analyse durchzuführen, um den Einfluss der Variablen und Ebenen auf die Ergebnisse eines Täterarbeitsprogramms festzustellen. Auf einer höheren Ebene können ExpertInnen die Reaktionen von Gerichten *im Allgemeinen* feststellen und diese Information wiederum als Variable in eine Datenmatrix eintragen. In einer derartigen Studie werden die relevanten Einflüsse empirisch ermittelt.

Um relevante Komponenten zu ermitteln, kann die Entwicklung für ein Verfahren zum Vergleich von Kontexten an Vergleichsstudien über rechtliche und unterstützende Maßnahmen ansetzen, z.B. EIGE (2012). Es gab auch bereits Projekte, die sich zum Ziel setzten, Kontexte in Bezug auf Maßnahmen für Opfer in verschiedenen europäischen Ländern abzubilden und zu vergleichen (z.B. EUCPN, 2013). Diese Arbeit zeigt, welche rechtlichen und praktischen Maßnahmen in den einzelnen Ländern umgesetzt werden.

Ein weiteres Beispiel für einen potentiell nützlichen Indikator ist der Gender Equality Index (GIA), der im Rahmen von EIGE (2012) entwickelt wurde. Obwohl Gewalt aufgrund eines Mangels vergleichbarer Daten nicht Teil des GIA ist, kann der GIA dennoch als übergeordneter Index für die Position eines Landes in Bezug auf Geschlechtergerechtigkeit verwendet werden; er kann auch als unabhängige Variable oder Kontrollvariable in eine Gleichung eingegeben werden.

Ein weiterer interessanter Zugang wurde von Hester und Lilley (2014) vorgeschlagen, in dem sie auf die Notwendigkeit hinweisen, verschiedene Risikofaktoren für Gewalt zu unterbrechen, und zwar auf vielfältigen Ebenen innerhalb eines ökologischen Modells (Gesellschaft, Institutionen, Familie, Peers, Einzelperson). Dieser Vorschlag fußt auf einem interaktiven Model von Risikofaktoren, das von Hagemann-White und KollegInnen (vgl.

---

<sup>6</sup> Oder eine andere Reihe von Alternativen zwischen 0 und 1.

Europäische Kommission, 2010) entwickelt wurde. Ihr Modell kann verwendet werden, um Wege zur Prävention von Gewalt in Beziehungen zu ermitteln. Die relevanten Präventivkomponenten können operationalisiert und als Variablen einer übergreifenden Analyse verwendet werden.

#### *4.5.3. Einrichtungsübergreifende Studien in einem einzigen Kontext*

Diese Art von Studien konzentriert sich darauf, *verschiedene* Täterarbeitsprogramme bei häuslicher Gewalt *innerhalb eines einzigen Kontexts* zu vergleichen. Es ist schwer feststellbar, welche Faktoren zu unterschiedlichen Ergebnissen führen, wenn man *verschiedene* Programme in *verschiedenen* Kontexten untersucht – daher wird empfohlen, den Kontext so konstant wie möglich zu halten, z.B. indem man zwei Arten von Programmen einer Region vergleicht.

Dieser Zugang geht nicht nur auf die verschiedenen Arten von Täterarbeitsprogrammen ein (z.B. eher kognitive Verhaltenstherapie oder eher psychodynamisch; oder mit und ohne zusätzliche Betreuung wie z.B. Alkoholentwöhnung oder Psychotherapie), sondern es könnten auch unterschiedliche Zuweisungswege verglichen werden, z.B. kognitiv behaviorale Programme einer Region, wo ein Programm mit Klienten arbeitet, die vom Gericht zugewiesen werden, mit einem andere Programm, das nur mit Selbstmeldern arbeitet usw. Dabei sollte man sich nicht nur fragen, welches Programm besser funktioniert als das andere, sondern auch, welcher Zugang für welche Klienten funktioniert, was bedeutet, dass man in der Aufnahmephase die Merkmale der Klienten untersuchen muss. Die Frage der unterschiedlichen Behandlungen für unterschiedliche Gruppen von Tätern kann bei Studien gestellt werden, die verschiedene Programme (innerhalb eines Kontexts) vergleichen und auch die Frage nach der Behandlungsintensität stellen können. Übliche Programme können den Bezugspunkt darstellen, mit dem Programme mit neuen Elementen verglichen werden.

Es ist wiederum wesentlich, in allen Studien dieselbe Methodologie zu verwenden. Außerdem wird die zum Zeitpunkt T0 erfasste Information über die Männer in den Programmen in höchstem Maße relevant sein. *Einrichtungsübergreifende Studien in einem einzigen Kontext* sind der richtige Zugang, um mehr über die Wechselwirkung zwischen Settings, Methoden, Gruppen von Tätern etc. herauszufinden. Daraus sollen sich Rückschlüsse auf differentielle Effekte verschiedener Arten von Täterarbeitsprogrammen bei häuslicher Gewalt ergeben.

#### *4.5.4. Studien, die wie unter Laborbedingungen durchgeführt werden*

Schlussendlich gibt es eine Art von Evaluierungsstudien, bei der es weder sinnvoll noch möglich ist (z.B. aus gesetzlichen Gründen), die Perspektiven der Opfer miteinzubeziehen (z.B. im Kontext von Täterarbeitsprogrammen in spanischen Gefängnissen). Andererseits wurden im Rahmen von Workstream 2 die einzigen kontrolliert randomisierten Untersuchungen in geschlossenen Institutionen wie Gefängnissen und psychiatrischen Kliniken vorgefunden (Hester et al., 2014). Gefängnisse bieten ein Setting, in dem es nicht

möglich ist, sofort die Wirkung von Täterarbeitsprogrammen festzustellen, weil die Insassen dort nicht mit ihren Partnerinnen zusammen leben und überhaupt nur eingeschränkten Kontakt zu ihnen haben. Dennoch bereiten Gefängnisprogramme als Teil von Resozialisierungsmaßen die Täter auf ein Leben nach der Haft vor und müssen daher als bedeutsame und sinnvolle Interventionen erachtet werden.

In Bezug auf die Evaluierung von Täterarbeitsprogrammen bei häuslicher Gewalt kann die Rolle von Gefängnisprogrammen mit der Laborsituation psychologischer Experimente verglichen werden. In Gefängnissen sind RKS grundsätzlich möglich. Vor allem Formen, in denen es Wartelisten gibt (d.h. Personen, die keine Behandlung erhalten und damit als Kontrollgruppe definiert werden und zu einem späteren Zeitpunkt eine Behandlung erhalten) sind brauchbar, um RKS in Gefängnissen ohne größere ethische Bedenken umzusetzen. Daher können Gefängnisprogramme dazu beitragen, Programme zu verbessern, indem sie sich auf verschiedene Behandlungskomponenten konzentrieren und sie systematisch untersuchen. Effektive Komponenten können dann in Programmen auch außerhalb von Gefängnissen umgesetzt werden (und auch durch *einrichtungsübergreifende Studien in einem einzigen Kontext* geprüft und validiert werden).

Von Studien, die wie unter Laborbedingungen durchgeführt werden, kann man differenzierte Aussagen über die Wirkung von Behandlung in Gefängnissen erwarten (vor allem auf der Grundlage der Perspektive des Mannes) und auch Aussagen über die Effektivität einzelner Behandlungskomponenten. Korrelationsstudien mit Rückfallsraten können durchgeführt werden, allerdings hängen Rückfälle von vielen Bedingungen ab, nicht nur vom Lernergebnis eines Mannes in einem Gefängnisprogramm. Daher sollten die Erwartungen in Bezug auf diese Art von Bewertung nicht zu hoch sein.

## 5. Zusammenfassung und weitere Schritte

Das ursprüngliche Konzept einer länderübergreifenden Evaluierungsstudie, die die europäische Diversität von Zugängen zur Täterarbeit nützt und eine Reihe von Fragen mit einem Mal beantwortet, musste modifiziert werden. Wir haben uns aus unterschiedlichen Gründen auf Täterarbeitsprogramme bei häuslicher Gewalt und koordinierte Zugänge konzentriert, auch weil der Rat der ExpertInnen aus Forschung und Praxis während der Workshops dieses Projekts in diese Richtung geführt hat. Wir haben die Grundlagen einer Methodologie in diesem Bereich umrissen. Diese Methodologie muss aber noch im Detail entwickelt werden, um systematischere und harmonisiertere Verfahren für die Evaluierung von Täterarbeitsprogrammen bei häuslicher Gewalt, sowohl auf nationaler als auch internationaler Ebene, zu erreichen. Die neue internationale Organisation WWP-EN kann und soll hier eine zentrale Rolle spielen, weil sie Standards für die Evaluierung von Täterarbeitsprogrammen unter ihren Mitgliedern entwickeln und verbreiten kann und weil viele zukünftige Evaluierungsprojekte mit diesen Mitgliedern zusammenarbeiten werden. Ein wichtiger Schritt, der jetzt gesetzt werden muss, ist es, eine Gruppe interessierter ForscherInnen und Fachleute innerhalb des WWP-EN zu gründen, die sich für diese Aufgaben verantwortlich fühlen. Bei der Entwicklung von Methoden und Standards können sie von den Ergebnissen des IMPACT Projekts und vor allem von den Empfehlungen in Workstream 3 ausgehen.

Es gibt einige Besonderheiten bei der Evaluierung von Täterarbeitsprogrammen bei häuslicher Gewalt. Eine der wichtigsten ist die Erkenntnis, dass die zuverlässigste Information über relevante Veränderungen von den Partnerinnen und Kindern der Täter kommt. Folglich ist eine Kooperation mit Opferschutzorganisationen nötig, sowohl auf der konkreten Ebene einzelner Einrichtungen als auch auf der Netzwerkebene mit Dachorganisationen wie WAVE.

Es hat sich herausgestellt, dass ein systemischer Zugang der geeignetste ist; daher sollten wir nicht anstreben, Täterarbeitsprogramme isoliert zu evaluieren, sondern das gesamte Interventionssystem sollte betrachtet werden. Vorgangsweisen dafür wurden in diesem Working Paper umrissen, indem neue Entwicklungen und Beispiele bewährter Evaluierungen von Täterarbeitsprogrammen bei häuslicher Gewalt übernommen wurden. Wir schlagen eine Struktur für Evaluierungsstudien vor, womit einzelne Studien miteinander verknüpft und verglichen werden können. Damit können verschiedene Designs in ein sinnvolles System gebracht werden, das auf den Entwicklungen der anderen Workstreams des IMPACT Projekts, früheren Studien und einigen vielversprechenden Entwicklungen bei der Evaluierung von Täterarbeitsprogrammen basiert. Wir hoffen, dass dieser Vorschlag dazu beitragen wird, die momentan fragmentierte und schwierige Situation der Vergleichbarkeit von Evaluierungen von Täterarbeitsprogrammen bei häuslicher Gewalt zu lösen. Schritt für Schritt sollten wir dazu in der Lage sein, relevante Effektgrößen optimierter Interventionen („Wer braucht was“), sowohl in Bezug auf Täter als auch auf die Opfer, herauszufinden und aufzuzeigen. Zuverlässige Effektgrößen für Täterarbeitsprogramme bei häuslicher Gewalt sind auch die Grundlage für eine sinnvolle Kosten-Nutzen- und Kosten-Effizienz-Analyse; üblicherweise zeigen derartige Analysen auf, dass es nicht nur eine ethische Frage ist, ob Mittel für Interventionen bereitgestellt werden, sondern auch, dass sie tatsächlich öffentliche Ressourcen einsparen helfen. Es ist ein häufiges Ergebnis, dass die Kosten für Interventionen gering ausfallen, im Vergleich zu den Kosten, die entstehen, wenn nicht gehandelt wird.

Weitere wichtige Aspekte und Überlegungen werden in separaten Beiträgen veröffentlicht, die auf den Aktivitäten des Workstream 3 des IMPACT Projektes basieren.

## 6. Literatur

- Alderson, S., Kelly, L. and Westmarland, N. (2013). Domestic violence perpetrator programmes and children and young people. London and Durham: London Metropolitan University and 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.
- Council of Europe (2011). Council of Europe Convention on preventing and combating violence against women and domestic violence. Retrieved from <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. and Westmarland, N. (2014) Ethics in violence and abuse research - a positive empowerment approach. *Sociological Research Online*, 19 (1) 2.
- EIGE (2013). Gender Equality Index. Report. Vilnius: European Institute for Gender Equality. Retrieved from <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.
- European Commission (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. Retrieved from [http://wwp.webseiten.cc/fileadmin/WWP\\_Network/redakteure/IMPACT/Daphne%20III%20Impact%20Working%20Paper%201%20rev.pdf](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: Council of Europe.
- Hester, M., Lilley, S.-J., O'Prey, L. & Budde, J. (2014). Overview and analysis of research studies evaluating European perpetrator programmes. Retrieved from [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](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. Retrieved from [http://wwp.webseiten.cc/fileadmin/WWP\\_Network/redakteure/Reports\\_2013/National\\_report\\_Hungary.pdf](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 and Durham: London Metropolitan University and 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.

Shadish, 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) (with 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*. Retrieved from [http://www.who.int/violence\\_injury\\_prevention/violence/world\\_report/en/summary\\_en.pdf](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*. Retrieved from <http://www.work-with-perpetrators.eu/index.php?id=86>