



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

Possibilités pour une étude européenne multi-sites et multi-pays de l'évaluation des programmes d'aide aux auteurs de violences conjugales.

Document de travail 3 du projet Daphne III "IMPACT: Évaluation des programmes européens d'aide aux auteurs de violences "

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.

Christian Scambor, Katarzyna Wojnicka & Elli Scambor



SOMMAIRE

1. Introduction.....	3
2. Méthodes	5
2.1 Analyse de la documentation.....	5
2.2. Application des résultats du Chantier 1	6
2.2.1. Aspects importants du Chantier 1	7
2.2.2. Conclusion	9
2.3. Application des résultats du Chantier 2	10
2.3.1. Aspects importants du Chantier 2	10
2.3.2. Conclusion	12
2.4. Entrevues avec les experts et les groupes de contrôle	13
3. Résultats: Points de vue des experts	14
3.1. Questions	15
3.2. Possibilités	15
3.3. Avantages	16
3.4. Défis	17
3.5. Obstacles	18
3.6. Résultats sortis des discussions avec les experts et les groupes de contrôle	19
4. Comment avancer avec ces évaluations des PTAV	20
4.1. Point de départ : où en sommes-nous aujourd’hui	20
4.2. La direction à prendre	21
4.3. Critères des projets d’évaluation	22
4.4. Promotion des critères	23
4.5. Une approche exhaustive vers l’évaluation des PTAV et le rôle des études multi- pays.....	24
4.5.1. Études basiques à contexte unique	24
4.5.2. Études multi-sites à contextes multiples	28
4.5.3. Études multi-sites à contexte unique	31
4.5.4. Études en conditions de pseudo-laboratoire	31
5. Conclusion et prochaines étapes	32
6. Bibliographie.....	33

Liste des abréviations fréquentes

TCC ... Thérapie cognitive et comportementale

RCC ... Réponse communautaire coordonnée

PTAV ... Programme de Thérapie pour Auteurs de Violences Conjugales

ERC... Études Randomisées Contrôlées

T0 ... Moment donné pour l'évaluation: Au moment du référé; durant la phase d'admission

T1 ... Moment donné pour l'évaluation: "Pré", au début de l'intervention (p.ex. travail de groupe)

T2 ... Moment donné pour l'évaluation: "Pendant", pendant l'intervention (p.ex. travail de groupe)

T3 ... Moment donné pour l'évaluation: "Post", à la fin de l'intervention (p.ex. travail de groupe)

T4 ... Moment donné pour l'évaluation: "Suivi", une période de temps après la fin de l'intervention (p.ex. 6 mois après la fin du travail de groupe)

IMPACT ... projet Daphne III " Évaluation des programmes européens d'aide aux auteurs d'agressions "

WAVE ... Women Against Violence in Europe (organisation internationale)

WWP ... Work with perpetrators (assistance aux auteurs de violences conjugales et leurs victimes)

WWP – EN ... Work With Perpetrators – European Network (Réseau Européen - organisation internationale)

1. Introduction

Selon les experts en la matière, l'évaluation des programmes d'aide aux auteurs de violences conjugales est essentielle non seulement pour les professionnels et les membres du personnel des organisations offrant ces programmes car elle leur permet de contrôler et d'améliorer la qualité de leur travail, mais aussi pour aider les responsables politiques à prendre des décisions stratégiques (Geldschläger et al., 2014; Hester, Lilley, O'Prey, Budde, 2014). Malheureusement, à ce jour, il n'y a eu que quelques publications de ces évaluations dans certains magazines scientifiques – malgré le fait que, lors de leur revue du Chantier 2, Hester et al. (2014) ont trouvé au moins 65 études nationales (publiées dans des revues scientifiques ou en « littérature grise »). Pourtant, ces études semblent fragmentaires. La majorité des projets ont été réalisés à un seul endroit et seulement quelques-uns d'entre eux pouvaient être décrits comme projets d'évaluation multi-sites. En général, ces rapports consistant en des études nationales, ont des approches différentes, différentes théories du changement, méthodes, outils etc. Les résultats du Chantier 2 donnent l'impression que le paysage de l'évaluation des programmes européens d'aide aux auteurs de violences est fractionné et complexe et les conclusions méta-analytiques sur le travail avec les auteurs de violences sont difficiles à tirer. Cette situation résulte du fait que les spécificités européennes ne sont pas vraiment prises en considération dans les débats internationaux sur les études d'évaluation des programmes d'aides aux auteurs de violences. Cependant, il faut dire que « ... les données existantes (provenant d'études faites en Amérique du Nord par exemple) ne peuvent pas facilement se généraliser à un contexte européen » (Hester et al., 2014, p. 3).

Ainsi, les buts principaux du projet « IMPACT: Évaluation des programmes européens d'aide aux auteurs de violences », financé par la Commission Européenne (Programme Daphne III), ont été définis comme suit :

- a) Combler le manque de connaissances concernant les évaluations des programmes d'aide aux auteurs de violences en Europe, et
- b) Offrir des solutions pour harmoniser le contrôle des résultats (une première étape vers une harmonisation d'une méthodologie d'évaluation)

Cela se fera :

- en fournissant une vue d'ensemble et une analyse des pratiques actuelles de contrôle des résultats des programmes d'aide aux auteurs de violences en Europe (Chantier 1) ;
- en fournissant une vue d'ensemble et une analyse des études de recherche évaluant les programmes d'aide aux auteurs de violences (Chantier 2) ;
- en identifiant les possibilités et les obstacles à une étude de recherche sur les résultats provenant de plusieurs pays Européen. (Chantier 3) ;
- en développant un guide et des directives de bonnes pratiques pour la mesure des résultats (Chantier 4).

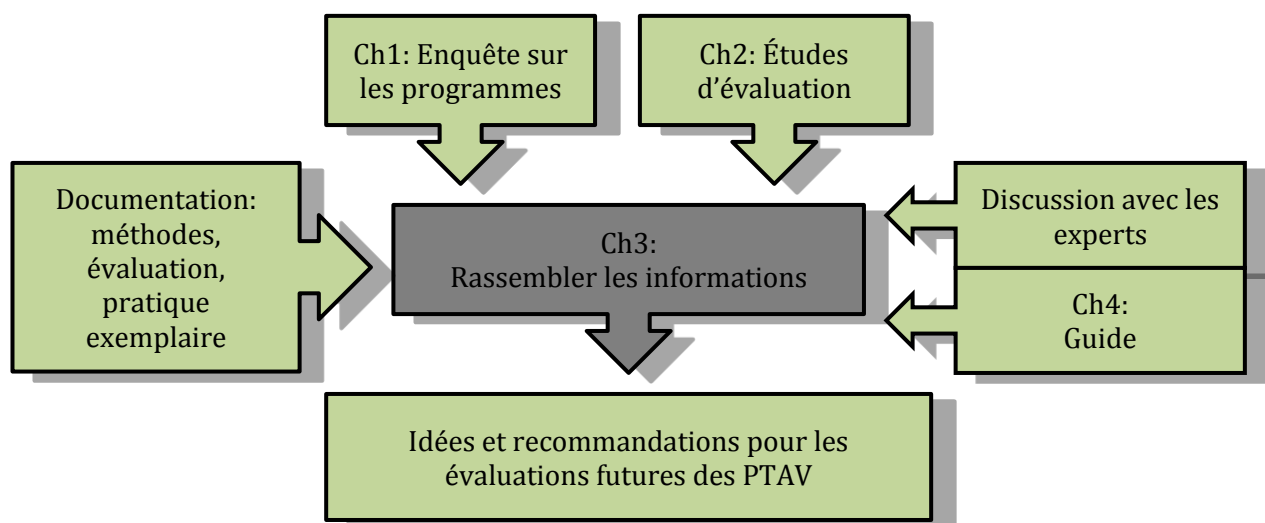
Ce document de travail présente la méthodologie et les résultats des activités exercées dans le Chantier 3. L'identification des possibilités et des obstacles à une étude de recherche sur les résultats provenant de plusieurs sites et de plusieurs pays Européen s'est basée sur :

- les résultats (intermédiaires et finaux) du Chantier 1: vue d'ensemble des pratiques de contrôle des résultats (sondage) ;

- les résultats (intermédiaires et finaux) du Chantier 2: vue d'ensemble des études européennes d'évaluation ;
- la revue documentaire dans le domaine des méthodologies de recherche et d'évaluation ;
- les interviews et discussions menées par des experts et professionnels travaillant dans le domaine de l'aide aux auteurs de violences et les services coopérants.

En vue des résultats de l'enquête du Chantier 1, on peut se demander dans quelle mesure les méthodes et les outils de comparaison pour mesurer ou évaluer les résultats sont déjà mis en place par les différentes organisations, et quelles conditions nationales ou institutionnelles requises peut-on identifier comme des avantages ou des obstacles aux projets d'évaluation multinationaux. S'appuyant sur les connaissances acquises dans la révision des recherches du Chantier 2, on a rassemblé des éléments méthodologiques pour l'évaluation des projets montrant une forte probabilité de transfert à différentes spécificités nationales. Dans ce sens, le Chantier 3 traite de la faisabilité des projets d'évaluation qui comprennent plus d'un site et plus d'un pays. Les activités du Chantier 3 ont été liées au Chantier 4 afin d'informer le développement du produit « guide » pour que ce produit contribue à l'harmonisation des pratiques relatives à la mesure des résultats des programmes et ainsi favorise une mise en œuvre des projets d'évaluation multi-pays.

Figure 1. Position et fonction du Chantier 3 dans le cadre du projet IMPACT.



Note. Ch... Chantier. PTAV... Programmes de Thérapie pour Auteurs de Violence conjugale

Ce document de travail résume le travail entrepris dans le cadre du chantier 3 en se focalisant particulièrement sur les apports spécifiques apportés par les experts et sur nos conclusions concernant les futurs projets d'évaluation dans le domaine du travail avec les auteurs de violences. Une analyse plus détaillée des questions évoquées dans ce document sera publiée dans 2 autres articles scientifiques.

Les auteurs désirent remercier tous les partenaires et les associés au projet IMPACT, tout particulièrement Heinrich Geldschläger et Marianne Hester et leurs équipes, pour leur avoir fourni des résultats intermédiaires importants qui les ont aidés à faire avancer le Chantier 3. Nous voudrions également remercier tous les experts qui ont pris part aux interviews et aux débats.

2. Méthodes

Le but du Chantier 3 était d'identifier les possibilités et les obstacles à des projets d'évaluation multi-site et multi-pays des programmes dans le domaine du travail avec les hommes auteurs de violences envers les enfants et entre partenaires intimes (PTAV – programme de thérapie pour les auteurs de violences conjugales). L'approche que nous avons prise pour identifier les modèles d'évaluation qui seraient applicables à différents pays européens et contextes consistait de :

- (1) l'analyse de documents existants sur l'évaluation des PTAV (effectuée en Europe et ailleurs) ainsi que d'autres documents pertinents,
- (2) l'utilisation des résultats des Chantier 1 et 2 dont les données ont été analysées de façon plus détaillée, et
- (3) d'interviews et de débats menés avec des experts et praticiens internationaux travaillant avec les auteurs de violences conjugales.

Il convient de souligner que l'objet principal du Chantier 3 concernait (a) *les résultats* (et non les procédés et autres moyens d'évaluer les programmes) et (b) *l'évaluation* et par conséquent les projets *de recherche*, qui vont au-delà de la mesure des résultats du travail au jour-le-jour des programmes. (Néanmoins, ces aspects sont liés et souvent considérés comme un tout : les résultats dépendent des procédés sous-jacents ; et idéalement, les méthodes utilisées pour la mesure des résultats sont compatibles ou même font partie de la méthodologie utilisée pour l'évaluation des projets).

2.1 Analyse de la documentation

Les informations récoltées parmi ces publications ont permis la préparation d'une vue d'ensemble interne des modèles d'évaluation les plus communs, les outils, les stratégies et les méthodes, et ont permis de formuler les questions destinées aux interviews des groupes de contrôle et des débats d'experts.

Entre autres, les importantes sources d'informations ont été :

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); ainsi que deux documents de travail du projet IMPACT: *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).

Une autre source importante d'informations pour le Chantier 3 a été la littérature de méthodologie sur l'évaluation (p.ex. Bamberger, Rugh & Mabry, 2006; Rossi, Lipsey & Freeman, 2004; Shaddish, Cook & Campbell, 2002; Stufflebeam & Shinkfield, 2007). Les termes spécifiques et les glossaires de ces documents ont été communiqués aux groupes du projet afin de faciliter l'usage d'une terminologie commune tout au long du projet.

Parallèlement au projet IMPACT, un projet d'évaluation d'importance considérable, sous le nom de MIRABAL¹, a été réalisé au Royaume-Uni. Ce projet a été établi pour répondre à des questions qui pouvaient également s'appliquer au projet IMPACT. Le projet partenaire Respect (R-U.) joue un rôle déterminant dans le projet MIRABAL, et ainsi, le transfert d'importants développements entre les deux projets a été possible. Les contacts et les discussions entre les deux projets ont été une source précieuse d'information (p.ex. au séminaire à Bristol (R-U.) en 2013, où Nicole Westmarland a présenté les développements récents et la méthodologie du projet MIRABAL). Les publications effectuées à ce jour et pertinentes au projet MIRABAL ont été une source inestimable d'information (Alderson, Kelly & Westmarland, 2013; Alderson, Westmarland & Kelly, 2012; Downes, Kelly & Westmarland, 2014; Phillips, Kelly & Westmarland, 2013; Westmarland & Kelly, 2012; Westmarland, Kelly & Chalder-Mills, 2010). Les résultats du projet MIRABAL seront présentés début 2015 et ils donneront sans doute une nouvelle impulsion pour les projets d'évaluation multi-sites dans le futur. Il importe de remarquer que l'expert américain Edward Gondolf (2012) a donné le projet MIRABAL en exemple pour une nouvelle tendance dans la recherche sur l'évaluation, qu'il appelle « évaluations menées par les praticiens » (p.81) et qui ont pour objectif de palier aux insuffisances des approches académiques parfois inappropriées à l'évaluation des programmes d'aide aux auteurs de violences.

2.2. Application des résultats du chantier 1

Le Chantier 3 s'intéressait particulièrement aux modèles d'évaluation, par exemple les concepts, les outils et les procédures d'évaluation. Dans le cadre du Chantier 1, une enquête a été réalisée afin de récolter des informations concernant les programmes européens d'aide aux auteurs de violences domestiques ainsi que leurs procédés pour mesurer leurs résultats et évaluer les performances de leur travail. Par conséquent, nous espérons obtenir des informations sur des pratiques prometteuses qui ont déjà été mises en œuvre ou déjà utilisées dans des projets d'évaluation, et cela montrerait un certain degré de transférabilité. Bien que les auteurs affirment que « ... l'accent n'est pas sur une étude scientifique des résultats mais sur le contrôle au jour-le-jour des résultats que les programmes d'aides aux auteurs de violences effectuent régulièrement comme partie intégrante de leurs prestations de service » (Geldschläger et al., 2014, p. 4), nous espérons obtenir des indications du Chantier 1 afin de planifier judicieusement les études d'évaluation futures. En examinant les résultats du Chantier 1, nous voulions explorer la possibilité de l'existence d'un dénominateur commun entre les éléments déjà en place et des approches qui pourraient être utilisées lors d'études ultérieures.

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

2.2.1. Aspects importants du Chantier 1

Lors de l'enquête du Chantier 1, les données de 134 programmes provenant de 22 pays ont été recueillies, et les résultats publiés dans un document de travail (Geldschläger et al., 2014). Les résultats les plus importants et leurs implications pour le Chantier 3 sont résumés ci-dessous. Pour plus de détails, le lecteur est renvoyé à Geldschläger et al. (2014).

Cas de référé

La plupart des programmes d'aide aux auteurs de violences travaillent aussi bien avec les hommes qui se sont portés volontaires qu'avec des hommes envoyés par une agence. Un nombre de programmes excluent certains cas de référé ou se concentrent sur certains cas uniquement (22% des programmes ne travaillent qu'avec des hommes adressés par le tribunal alors que 19% ne travaillent pas du tout avec ces genres de cas ; Geldschläger et al., 2014, p. 12). Beaucoup de programmes privilégient le travail avec les hommes envoyés par le tribunal ou ceux portés volontaires (respectivement référés par des agences externes) – mais la majorité des programmes travaillent avec une grande variété des cas. Les études d'évaluation futures multi-sites devraient tenir compte de cette variété.

Coopération

Presque tous les programmes (98%) dans l'enquête collaborent avec un réseau de partenaires dont les plus importants sont : les services sociaux, la police, les services de protection des victimes, les tribunaux, les services de traitement de l'alcoolisme et la toxicomanie. 75% des programmes disent faire partie d'une alliance interinstitutionnelle contre les violences domestiques, et un certain nombre des programmes offrent un service de soutien aux victimes au sein de leur organisation. Nous pouvons donc conclure que les programmes de l'enquête suivent une « approche systémique » à un niveau relativement élevé, dans une variété de pays différents. La nature et la qualité des réseaux peuvent varier d'un pays à pays et cela doit être pris en compte lors de la conception d'évaluations multi-sites. Cependant, ces conclusions signifient également qu'il est possible de se concentrer en premier sur les programmes qui font partie d'un réseau coordonné d'institutions et de services. Les programmes qui diffèrent de cette approche offrent de l'intérêt – ils pourraient faire partie d'un programme d'évaluation qui pourrait être étudié lors d'une seconde phase, étant donné qu'ils requièrent des concepts et des comparaisons différents.

Admissions

Presque tous les programmes (95%) ont, d'une manière ou d'une autre, une procédure d'admission. Cette phase d'admission peut aller d'un contact initial à une phase plus longue de consultations individuelles ou d'évaluations. Cela est en fait un aspect très utile pour la réalisation de projets d'évaluation. Pour beaucoup de programmes, la mise en œuvre d'une forme de collecte de données au début est une procédure « normale ». Deux tiers des programmes utilisent certains outils à l'admission – bien qu'il n'y ait pas d'outil prévalent. Une variété de questionnaires, de manuels et autres outils sont nommés. Une harmonisation est donc nécessaire à ce niveau-là, mais il reste important que la plupart des projets s'habituent, par principe, à travailler avec des outils. Les futures études d'évaluation peuvent s'appuyer sur ces compétences.

La majorité des programmes peuvent appliquer des critères d'inclusion ou d'exclusion, dans

le cadre de leur procédure d'admission (p.ex. trois quarts d'entre eux stipulent « l'abus d'alcool et de drogues » et « les troubles mentaux » comme des conditions d'exclusion). Cela signifie que la majorité des programmes peuvent commenter sur certaines caractéristiques concernant leurs participants, que ce soit des caractéristiques comportementales ou psychologiques.

À nouveau, il s'agit est d'un résultat important pour les études multi-sites d'évaluation : il semble possible de mettre en œuvre une procédure de collecte de données avant le début du programme, afin d'étudier les différences entre certains types de participants et, respectivement, les abandons qui surviendraient plus tard.

Configuration

Environ deux tiers des programmes utilisent le travail de groupe (pour être plus tard combiné à d'autres méthodes d'intervention). Les consultations individuelles sont également fréquentes mais le travail de groupe reste la méthode d'intervention prédominante. Pour les études multi-sites d'évaluation, les deux approches (travail de groupe et consultations individuelles) semblent intéressantes. D'autres approches sont moins fréquentes (p.ex. conseils matrimoniaux, médiation). Ces approches peuvent être incluses dans certains concepts d'évaluation, mais elles demanderaient des modèles d'évaluation bien spécifiques. En ce qui concerne les « doses » du traitement, il y a une large variation en fonction des programmes, dont la plupart offrent entre 14 et 52 sessions sur une durée de 14 à 52 semaines. Cette variation est importante pour les études d'évaluation en termes d'approcher dose-effet, ce qui veut dire que des interventions à long-terme peuvent être comparées aux interventions à court-terme.

Approche de travail

Concernant l'approche des programmes pour leur travail concret fait avec les auteurs de violences, la plupart se placent dans le domaine de la Thérapie Cognitive et Comportementale (TCC) et de l'approche psycho-éducative. En combinant la TCC, l'approche psycho-éducative et l'approche basée sur la méthode Duluth, deux tiers à trois quarts des programmes se basent sur ces méthodes. Ces approches ne sont, bien sûr, pas identiques mais elles ont certains éléments en commun, particulièrement à un niveau théorique (théorie de l'apprentissage). Aux Etats-Unis, Gondolf (2012) mentionne deux catégories d'approche dans le travail avec les auteurs de violences : les approches c-comportementales et psycho-dynamiques (respectivement les approches « nouvelle psychologie »), qui sont par elles-mêmes deux approches bien diversifiées. Dans le cas de l'Europe, nous voyons une variété de méthodes, parfois combinant des éléments de différentes approches.

La mesure des résultats

En ce qui concerne la mesure des résultats, 81% des programmes affirment mesurer leurs résultats d'une façon ou d'une autre (p. 33). La source d'information la plus importante est les participants eux-mêmes (94% de ceux affirmant avoir mesuré leurs résultats), les intervenants (63%) et les partenaires/ex-partenaires (60%), ainsi que d'autres services (54%), les rapports de police/justice (43%) et les services de soutien aux victimes (32%). Deux tiers des programmes affirment offrir un programme de suivi.

Ce résultat montre que l'intégration de sources d'information autres que les participants aux programmes est une pratique courante à travers les projets, fait sur lequel les futures études d'évaluation peuvent s'appuyer. Le *point de vue des partenaires* est considéré comme une

source d'information centrale dans les projets d'évaluation, et on peut trouver des programmes qui prennent en compte les informations provenant des partenaires/ex-partenaires dans la plupart des pays. D'après les informations recueillies dans l'enquête, cette pratique semble être plus courante dans certains pays, tels que le Royaume-Uni, que dans d'autres (p.ex. en France ou en Suisse). De plus, dans certains pays, les programmes sont fermement liés au système de Justice criminelle, c'est-à-dire que « ... comme attendu les programmes ordonnés par la Cour ont tendance à ignorer les informations venant des partenaires et des services de soutien aux victimes, car beaucoup de ces programmes ont stipulé qu'ils ne contactaient pas les partenaires... et ne mesurent pas les résultats venant de ceux-ci... » (Geldschläger et al. 2014, p. 46).

Néanmoins, nous trouvons des programmes dans 17 pays, sur tout le continent, qui prennent en compte les informations des partenaires/ex-partenaires pour mesurer les résultats. Nous pouvons ainsi conclure qu'il ne devrait pas y avoir de problèmes, en principe, à une étude multi-pays qui inclurait les partenaires dans l'évaluation des résultats des programmes. Là où ce n'est pas possible, les programmes respectifs joueront un rôle différent dans une future stratégie d'évaluation future.

En ce qui concerne les *hommes participant aux programmes*, les résultats les plus fréquents que l'on a pu mesurer étaient : la non-violence / baisse des incidents de violences, les attitudes et les convictions personnelles, l'aptitude des participants à communiquer et la baisse dans le risque de violences. Les aspects psychologiques, les compétences parentales et la qualité de vie ont été également mesurés à mi-parcours ou aux trois quarts de la durée des programmes.

Les aspects les plus importants couverts dans les questionnaires donnés aux partenaires/ex-partenaires étaient : les violences, les sentiments d'insécurité, la baisse du risque de violences, la sûreté des enfants, la qualité de vie, l'attitude de l'homme, l'aptitude de l'homme à communiquer, les compétences de l'homme en tant que père.

En tout, il y a une grande variabilité des méthodes de mesure des résultats des programmes d'aide aux auteurs de violences. « Une des conclusions principales de cette enquête est donc que les programmes d'aide aux auteurs de violences en Europe sont loin d'un consensus sur les méthodes standards pour la mesure des résultats et un besoin d'harmonisation semble très apparent » (Geldschläger et al., 2014, p. 56). Dans le Chantier 4 du projet IMPACT, nous faisons un pas dans cette direction.

2.2.2. Conclusion

La conclusion que nous pouvons tirer des résultats de l'enquête dans le Chantier 1 est qu'il n'y a pas de problèmes en principe pour mesurer les résultats dans plusieurs pays et pour plusieurs programmes de façon exhaustive, en incluant la perspective des hommes participant aux programmes, des partenaires et des enfants, des documents officiels, des intervenants. Pourtant, des ressources complémentaires sont nécessaires pour rejoindre un projet d'évaluation, pour la plupart des programmes². Les praticiens semblent

² Cela s'applique également pour la mesure des résultats des programmes d'aide aux auteurs de violences: pour ce qui est des obstacles à la mesure des résultats, les problèmes principaux pour les programmes sont les contraintes de temps et de ressources. Une proportion plus réduite de programmes (environ un tiers) voient « le manque de méthodologie » et « le contexte légal/institutionnel » comme des obstacles à la mesure de leurs résultats (pour obstacles légaux, voir ci-dessus : programmes liés au système judiciaire dans certains pays). Cependant, presque tous les programmes ayant participé à l'enquête désirent améliorer leurs méthodes de mesure de résultats et d'évaluation et participer à un échange international. Il y a ainsi un fort besoin d'augmenter les ressources humaines, économiques et méthodologiques et les ressources d'analyse de données, des directives et recommandations, formations et consultations.

ouverts à l'évaluation et différentes procédures pour collecter les données sont en place. Néanmoins, les études d'évaluation vont bien au-delà de la mesure et du contrôle des résultats. En conséquence, les études d'évaluation multi-sites devraient être exécutées comme des projets séparés bénéficiant d'un financement supplémentaire. Si nous nous basons sur les conclusions du Chantier 1, il semble inapproprié de mettre en place une étude d'évaluation multi site/multi-pays qui se repose sur les pratiques actuelles des programmes, que ce soit la mesure des résultats ou la collecte des informations durant la phase d'admission. Les pratiques existantes sont tout simplement trop hétérogènes pour être utiles à une étude dans un contexte multi-sites. Pour que tous ces programmes puissent participer à une étude d'évaluation, des procédures supplémentaires sont requises, et donc des ressources supplémentaires sont nécessaires pour la grande majorité de ces programmes.

2.3. Application des résultats du Chantier 2

Dans le Chantier 2, l'accent a été déplacé de la pratique vers la recherche. Le but était d'acquérir des connaissances sur les approches utilisées dans les études de recherche d'évaluation des programmes d'aide aux auteurs de violences en Europe. L'équipe de chercheurs a identifié des études publiées dans des revues scientifiques et dans la « littérature grise », en langue anglaise et autres langues européennes. Les méthodes, concepts, mesures de résultats et autres aspects de ces études ont été analysés. C'est pourquoi les conclusions du Chantier 2 étaient d'importance primordiale pour les objectifs du Chantier 3 (c.-à-d. offrir des recommandations pour les futures études d'évaluation concernant les modèles d'évaluation).

2.3.1. Aspects important du Chantier 2

Fréquence par pays

Hester et al. (2014) décrivent en détails leur approche pour identifier les études existantes. Quatre méta-analyses et synthèses ont été identifiées et étaient basées sur quinze études en total. Il est intéressant de noter que la plupart de ces études provenaient du Royaume-Uni (7) et d'Espagne (5), et que ces deux pays faisaient également partie des pays offrant le plus grand nombre des programmes identifiés lors du Chantier 1. Avec des étapes supplémentaires pour identifier les études d'évaluation, publiées ou non (ou publiées en tant que littérature grise), un total de 65 études ont pu être utilisées pour une analyse plus détaillées dans le Chantier 2. À nouveau, nous remarquons que beaucoup de ces études viennent d'Espagne et du Royaume-Uni, mais aussi de pays d'Europe centrale³. Nous pouvons constater l'existence d'un chevauchement des fréquences concernant *les programmes* (Chantier 1) et *les études d'évaluation* (Chantier 2) qui paraît tout à fait raisonnable : les pays offrant (ouvertement) le plus de programmes d'aide aux auteurs de violences ont tendance à effectuer et publier plus d'études d'évaluation.

³ Pour leurs analyses, Hester et al. (2014) ont classé les pays en 5 régions où on a identifié des études d'évaluation : Ouest (R-U, Irlande), Sud (Espagne, Portugal), Nord (Suède, Danemark, Finlande, Islande), Est

(Croatie), and Centre (Allemagne, Autriche, Suisse, Pays-Bas).

Problèmes liés aux études

Nous avons rencontré une série de problèmes dans l'analyse des études d'évaluation. Quelques-uns des plus importants sont mentionnés ci-dessous :

- *Conception*: en ce qui concerne la conception de la recherche, nous n'avons trouvé pratiquement aucune Étude Randomisée Contrôlée (ERC). Des n=58 études provenant de cinq régions européennes et qui adressaient la question des *résultats* (concentrées uniquement sur les résultats ou sur les résultats et les procédés), seulement 2 étaient basées sur un concept d'ERC, et celles-ci étaient effectuées dans des prisons où des cliniques psychiatriques.
- *Groupes de contrôle*: Beaucoup d'études manquaient de groupe de contrôle, ou avaient des groupes de contrôle trop petits. Ceci est limité grandement le volume des études disponibles puisque les groupes de contrôle (ou autre élément de contrôle équivalent permettant une inférence causale) sont cruciaux à la validité interne de toute étude.
- *Terminologie*: dans une étude, on peut trouver un éventail de définitions de termes cruciaux (p.ex. « début », « aboutissement »). L'usage uniforme de cette terminologie est inexistant.
- *Moments de l'observation*: il existe différentes approches et méthodologies concernant les moments ou les données doivent être collectées. Hester et al. (2014) ont défini un procédé pour leur analyse qui devrait être utilisé et promu pour obtenir une compréhension commune concernant les moments pour l'observation dans les études d'évaluation. Il consiste en 5 moments où des informations doivent être collectées pour une évaluation (T0 : avant le début ; T1 : au début ; T2 : durant l'intervention ; T3 : à la fin ; T4 : lors du suivi). Ce système est détaillé un peu plus bas.
- *Taux de perte*: les taux de perte étaient obscurs, ou la phase dans laquelle elles se déroulaient n'était claires. Les informations sur les tailles des échantillons et les taux de perte pour chaque étape du processus d'intervention étaient inexistantes dans presque chaque cas. Les informations données sur les échantillons et leur taille variaient énormément, dû au manque d'harmonisation et d'une terminologie commune établie pour les études d'évaluation du travail des programmes d'aide aux auteurs de violences.
- *Cas de référé*: Il est essentiel d'avoir des informations concernant les moyens par lesquels une personne peut être placée dans un programme (participation volontaire ou ordonné par la Cour, autres moyens de référé). De plus, les informations concernant les raisons d'abandon, données sociodémographiques et autres informations de tous les sous-groupes (participant ayant été référé avant le début de la thérapie ; au début de la thérapie ; abandons ; finissants) sont essentielles. Dans la majorité des cas, ces informations manquaient.

Mesurer les résultats et l'impact

En ce qui concerne les résultats, les données apportées par les auteurs de violences eux-mêmes, les données officielles venant du système de Justice et les données venant des victimes étaient utilisées dans toutes les régions, excepté la région Sud où les données venant des partenaires n'apparaissaient pas dans les études disponibles. Pourtant, il y avait

des programmes dans la région Sud qui avaient inclus les points de vue des victimes, comme nous l'avons trouvé dans l'analyse du Chantier 1. Cependant, il semble qu'il y ait des différences dans la pratique et les limitations légales des programmes basés en milieu carcéral et communautaire. Les programmes basés en milieu carcéral n'intègre pas les perspectives des partenaires, mais ils sont évalués plus souvent que ceux opérant en milieu communautaire. De plus, ils utilisent beaucoup d'outils psychométriques validés/standardisés, se concentrant sur les aspects psychologiques et les changements des hommes auteurs de violences. Cela reflète une forte approche psychologique dans l'évaluation des programmes d'aide aux auteurs de violences en milieu carcéral dans la région Sud, dont la plupart des études viennent d'Espagne.

Les mesures de résultats les plus complètes ont été effectuées dans la région Ouest, « ... car elles avaient pris en compte les changements d'attitude et des violences des hommes envers les femmes, les cas de récidive (rapports officiels et déclarations volontaires), la victimisation répétée, la qualité de vie, la sûreté et le bien-être des femmes/victimes et leurs enfants, le niveau de stress parental et leurs abus d'alcool/de drogues » (Hester et al., 2014, p.26). Des outils d'évaluation des risques peuvent être trouvés dans les études provenant des régions Ouest et Sud.

En général, les aspects étudiés et les outils utilisés varient largement d'une région à l'autre (et d'une étude à l'autre). Ce manque de méthodologie cohérente rend la comparaison d'études difficile voire impossible.

Ainsi, lorsque qu'il faut donner une conclusion globale sur l'efficacité des programmes d'aide aux auteurs de violences, en se basant sur les études d'évaluation disponibles, on ne peut fournir qu'une vue d'ensemble très générale: il y a soit des conclusions statistiquement significatives qui pointent dans la direction de changements positifs ; ou alors, on trouve des résultats prometteurs mais qui n'ont aucune pertinence du point de vue des statistiques. Cette tendance a été trouvée pour une grande variété de dimensions (changement comportemental, changement d'attitude, changement psychologique, changement dans les facteurs risques, motivation et fidélité). Il y a pourtant des limitations dans les conclusions concernant l'impact, tel que le nombre réduit des échantillons dans beaucoup de cas, les taux de perte élevés (ou les données inexistantes concernant les taux de pertes) et une dépendance trop importante aux informations provenant directement des hommes dans la région Sud.

2.3.2. Conclusion

En ce qui concerne le Chantier 3, et malgré les problèmes rencontrés dans les études, les résultats indiquent qu'il est possible d'obtenir et de combiner les données dans toutes les régions. Il y a différentes approches concernant l'évaluation, mais il semble possible d'appliquer certains outils et d'intégrer certaines sources d'information provenant de plusieurs pays (à l'exception de certains cas, par ex : les programmes basés dans les prisons qui jouent un rôle différent dans les futures stratégies de recherche sur l'évaluation).

Le problème principal aujourd'hui semble être l'incohérence des méthodologies utilisées pour évaluer les programmes d'aide aux auteurs de violences, qui rend la comparaison des études difficiles sur plusieurs plans (conception et groupes de contrôle, terminologie, outils, sources des données). Une harmonisation des méthodologies d'évaluation est requise afin de remédier au fractionnement des recherches sur l'évaluation des programmes d'aide aux auteurs de violences. Cela semble possible selon les conclusions tirées par les auteurs du

document de travail sur le Chantier 2 :

« Il y avait des différences et des similitudes intéressantes entre et à l'intérieur des régions, et il serait possible, dans l'ensemble, de prendre des éléments des différentes approches afin de commencer à développer une méthodologie d'évaluation robuste. » (Hester et al., 2014, p. 39)

2.4. Entretiens avec les experts et les groupes de contrôle

En ce qui concerne, les études d'évaluation des programmes d'aide aux auteurs de violences, de nombreuses parties prenantes peuvent être identifiées, chacune apportant leurs propres intérêts et leurs perspectives, qui peuvent différer ou bien chevaucher :

- *Les victimes-survivantes de violences domestiques*: les femmes partenaires et les enfants voudront savoir si le programme agit efficacement sur le comportement violent de leur partenaire/père et sur l'amélioration de leur sécurité.
- *Les professionnels du soutien aux victimes*: les professionnels dans le domaine du soutien aux victimes veulent connaître l'efficacité des programmes et si on peut justifier une augmentation des ressources financières qui pourraient autrement utiliser pour augmenter le soutien aux victimes.
- *Les intervenants dans la chaîne d'intervention* : la Police, les procureurs, les tribunaux, les agents de probation et autres parties prenantes sont également intéressées de voir si les PTAV ont un impact particulier (en plus des interventions déjà existantes)
- *Les intervenants dans les PTAV*: Ces professionnels seront intéressés de connaître les résultats de leur travail et voudront savoir particulièrement si leur travail est efficace ou non et si les violences diminuent et si c'est le cas comment peuvent-ils améliorer leur travail pour obtenir de meilleurs résultats.
- *Les directeurs des programmes de PTAV*: ils seront intéressés de connaître les effets des programmes ainsi que comment améliorer la qualité de ces programmes, par exemple leur intérêt est de démontrer l'efficacité et l'amélioration des résultats.
- *Le public*: De façon générale, le public pourrait être intéressé par une baisse des cas de violences conjugales dans la société et une augmentation de la sécurité. Les débats publics sur le sujet se passent souvent à travers les médias qui ont leur propre logique et leurs propres intérêts. Il se peut que les médias de qualité soulèvent des questions similaires à d'autres groupes intéressés, pendant que d'autres sections des médias utiliseront un contenu plus choquant, tape-à-l'œil et extraordinaire. Ces groupes-là, demandent souvent que les peines contre les auteurs de violences soient plus lourdes et les lois plus strictes, cependant il y a un intérêt certain dans les effets de méthodes alternatives comme les PTAV.
- *Donateurs, politiciens*: Les parties prenantes de cette catégorie voudront montrer au public leurs activités pour combattre les violences ainsi que d'assurer que les fonds affectés aux programmes ont des effets et mènent à l'aboutissement des objectifs prédéfinis. Parallèlement, ils désireraient optimiser l'usage des ressources, c'est-à-dire répondre aux besoins de rentabilité.

Lors des entretiens avec les groupes de contrôle et les discussions avec les experts (voir ci-dessous), la question des différences et similitudes dans les intérêts des différentes parties prenantes s'est trouvée problématique.

Les **entretiens avec les experts et les groupes de contrôle** (une méthode de recherche sociale qui combine les interviews d'experts et des groupes de contrôle ; voir Babbie, 2003; Silverman, 2013) ont été conduites en octobre 2013 pendant l'*Atelier A* du projet organisé à l'Université de Bristol, R-U. Quatre interviews de 45 minutes ont été conduites avec les groupes de contrôle et les 18 experts (chercheurs et intervenants travaillant avec les auteurs de violence) venant de 11 pays européens (Autriche, République Tchèque, Danemark, Finlande, Allemagne, Irlande, Pologne, Portugal, Norvège, Espagne et Royaume-Uni). Pour chaque groupe, un facilitateur (un membre de l'équipe travaillant sur le projet) a posé trois questions principales :

- À un niveau national ou local, quelles conditions favorables ou défavorables à l'élaboration d'un concept multi-sites dans différents pays voyez-vous ?
- À un niveau international, quelles conditions favorables ou défavorables à l'élaboration d'un concept multi-sites dans différents pays voyez-vous ?
- Quelles possibilités et obstacles, à votre avis, voyez-vous concernant l'idée suivante : « fournir un outil aux professionnels (dans le domaine du travail avec les auteurs de violences), pour recueillir des données comparables dans différents pays et centraliser ces données pour les analyser » ?

Après les entretiens, les facilitateurs ont préparé des rapports internes qui ont servi de base à la présentation donnée lors de l'*Atelier B* ainsi que lors d'analyses ultérieures (voir ci-dessous)

Les débats d'experts ont eu lieu durant l'*Atelier B* organisé en avril 2014 à Copenhague, Danemark. 41 professionnels, chercheurs et responsables politiques venant d'Autriche, de Bulgarie, de la République Tchèque, de Croatie, du Danemark, d'Estonie, de Finlande, d'Allemagne, d'Irlande, d'Italie, de Lituanie, des Pays-Bas, de Russie, d'Espagne, de Suède, de Suisse et du Royaume-Uni ont discuté des problèmes et questions soulevés dans la présentation *Possibilité d'Évaluation dans une Approche Multi Site / Multi-pays* :

- Quelles opportunités et obstacles je peux voir dans mon organisation ?
- Quelles questions doivent-être traitées dans une étude d'évaluation:
 - Est-ce-que cela fonctionne?
 - Dans quelle mesure / est-ce efficace?
 - Qu'est-ce qui fonctionne?
 - Questions spécifiques (qui récidive, quels sont les besoins etc.)

Les résultats de ces discussions conduites dans quatre groupes ont été recueillis et ont servi de base à des analyses ultérieures (voir ci-dessous). D'autres commentaires sur le concept d'une étude d'évaluation multi-sites/multi-pays des PTAV ont été rassemblés lors de la dernière conférence du projet, à Barcelone, Espagne, en octobre 2014.

3. Résultats: Points de vue des experts

Les résultats des entretiens avec les experts et les groupes de contrôle, ainsi que les résultats des débats d'experts ont été résumés dans une série de points et qui a été une source d'information pour savoir quelle direction prendre pour le développement d'une méthodologie pour l'évaluation des programmes d'aide aux auteurs de violences. Ces points ont été catégorisés en cinq groupes majeurs : les questions, les opportunités ; les avantages,

les défis et les obstacles.

3.1. Les questions

Une des questions les plus communes et les plus importantes qui est apparue au cours des entretiens et des discussions, faisait référence à la compréhension fondamentale des « programmes d'aide aux auteurs de violences ». Selon les experts de plusieurs pays, la définition de tels programmes peut varier : excepté les programmes typiques (par exemple les réunion de groupe régulières, menées par des facilitateurs professionnels, structurées, coordonnées et évaluées), les interventions suivantes ont été nommées comme des éléments intégrants ou satellitaires du travail avec les auteurs de violences envers partenaires intimes : numéro d'écoute, paramètres individuels, réunions informelles, volontaires, communautaires et/ou d'Eglise, réunions organisées par Alcooliques Anonymes etc. Ainsi, une des premières étapes dans l'élaboration d'une recherche européenne sur l'évaluation du travail avec les auteurs de violences serait de donner une **définition** commune mais également relativement large **des programmes d'aide aux auteurs de violences**.

Une autre question importante concernant le problème de **la similarité des programmes**. La plupart des experts partageaient l'opinion que seuls les résultats de programmes similaires (même approche, mêmes méthodes de travail, à caractère similaire, durée, taille etc.) devraient être comparés. Cette hypothèse a pourtant posé des doutes concernant le nombre (insuffisant) de programmes européens qui partageaient assez de critères pour pouvoir être comparés.

La question de la similarité a été suivie de la question du **type de programme** à étudier. Certains experts ont souligné le fait que les futures études devraient se concentrer sur un seul type uniquement : soit les programmes obligatoires, soit les programmes volontaires, tandis que d'autres experts proposaient d'étudier les deux approches ou de se concentrer sur les programmes « mixtes » (ceux comportant les deux de mode de référé).

Finalement, les experts ont abordé le problème de la **procédure de recherche** d'une future étude. Les questions concrètes ont prévalu, telles que qui seraient les chercheurs, qui devrait collecter les données etc. Les pour et les contres de différentes possibilités ont été débattus, par exemple faire appel à des chercheurs externes, indépendants, ou demander aux professionnels qui travaillent directement avec les auteurs de violences de recueillir les informations et peut-être procéder aux premières étapes d'analyse. La motivation et un intérêt certain dans l'évaluation du travail avec les auteurs de violences peut être un avantage dans ce dernier cas, mais les conflits d'intérêt peuvent présenter un inconvénient. Selon les experts, les deux procédés ont leur avantages et leurs inconvénients (différence dans l'accès aux données et aux échantillons, coûts, différences dans les niveaux de compétences analytiques, les problèmes de l'objectivité, contraintes temporelles etc.) et la décision concernant quelle procédure choisir devrait être aussi considérée.

3.2. Possibilités

Un grand nombre d'experts ont exprimé leur intérêt d'exécuter et de soutenir une forme d'étude de résultats multi-pays en Europe. Par conséquent, la question des possibilités d'une telle étude a été traitée en plus amples détails. Lors de discussions concernant cette

question, les **standards des programmes** ainsi que **la mesure des résultats** à être utilisés dans les études d'évaluation étaient des points particulièrement intéressants pour les experts.

Un point de départ dans la définition d'un standard minimum pour les programmes d'aide aux auteurs de violences et la mesure de leur résultats peut se trouver dans la *Convention pour la Prévention et la lutte contre la Violence à l'égard des Femmes et la Violence Domestique* (« Convention d'Istanbul », Conseil de l'Europe, 2011), dans les notes d'information sur les critères de résultat par Westmarland, Kelly et Chalder-Mills (2010) ou dans les *Directives pour le Développement des Standards pour les Programmes Travaillant avec les Hommes Auteurs de Violences Domestiques* (Work with Perpetrators [WWP]-Réseau Européen, 2008). Il faut également souligner que, selon les experts, ces standards devraient se baser sur une politique d'égalité de genre.

Le terme « minimum » a souvent été mentionné en relation avec les études de recherche et d'évaluation, dans le sens où un **minimum de standards** est nécessaire pour **chaque étude spécifique à un pays** ou pour chaque part de l'étude de façon à pouvoir « mettre en commun » les différents éléments. Plusieurs éléments qui ont déjà été utilisés à différents moments ont été identifiés, par exemple le modèle écologique, le modèle d'observation par étapes (pré-, post- et suivi), certains outils (p.ex. inventaires des violences et d'abus). Les chercheurs ont déjà commencé à décrire et analyser plusieurs problèmes relatifs, tels que le contexte national et européen (la législation, les moyens de référence etc.), les types d'auteurs de violences, la fréquence et le type de violences utilisées, les changements désirés (comportement violent, communication, opinions et identités, relations familiales etc.). Le débat a amené certains experts à premièrement soulever des questions de recherche (sans pour autant apporter de solutions au moment de la discussion) concernant la méthode à utiliser pour répondre aux problèmes suivants :

- Qu'est-ce qui contribue le plus à l'arrêt des violences et à l'amélioration de la sûreté des victimes ?
- Les hommes changent-ils? Pourquoi? Comment?
- Qu'est-ce qui fonctionne le mieux et pour qui?
- Quels (autres) facteurs affectent la vie des auteurs de violences (classe sociale, religion, les origines etc.)?

Finalement, les experts ont discuté des conditions qui pourraient améliorer la **qualité des futures études d'évaluation multi site/multi-pays**, p.ex. triangulation des méthodes de recherche (qualitatives, quantitatives, observations) triangulation des données et des théories, la prise en compte des perspectives des auteurs de violences, de leur victimes et des professionnels. De plus, selon les experts, l'implication d'organismes et d'ONG locales/spécialisées (groupes de migrants, groupes LGBT, groupes religieux) dans le processus de recherche peuvent également être hautement utiles.

3.3. Avantages

Durant tous les entretiens et les discussions, les experts ont souligné les avantages des études d'évaluation multi-sites. Dans une telle étude, il faudrait développer et utiliser une façon **efficace** et **utile** de mesurer les violences et les autres **mesures de résultats** qui ne seraient pas uniquement profitables aux professionnels de tous les pays européens mais contribueraient également à combler les lacunes dans les connaissances et la recherche sur les violences en général, car selon les experts, la méthodologie actuelle est insuffisante. Les

professionnels en particulier ont exprimé l'espoir que l'évaluation soutienne la **continuation des programmes existants**, puisque certains corps politiques associent l'apport de fonds à des résultats positifs de l'évaluation des programmes.

De plus, les experts ont signalé que l'idée d'intégrer les perspectives des victimes et des organisations de soutien aux victimes amènerait à **des études d'évaluation plus complètes**, y compris le **système plus large** qui était utilisé lors de l'intervention. Les études d'évaluation multi-pays en Europe pourraient également contribuer à la **promotion** de programmes d'aide aux auteurs de violences uniques et nationaux qui ne sont pas encore bien connus parmi les parties prenantes, et les résultats venant de ces programmes serviraient de **base de connaissances** pour les professionnels désireux d'améliorer la qualité de leur travail, éviter des erreurs communes etc. Par ailleurs, pour ajouter une dimension supplémentaire, l'attention portée sur une étude internationale sur le travail avec les auteurs de violences susciterait une **prise de conscience sociale** concernant l'existence et le rôle des programmes d'aide aux auteurs de violences dans certains pays, et, selon certains experts, cela aidera une promotion plus efficace à un niveau aussi bien national qu'international. Si une étude d'évaluation multi-pays en Europe était effectuée comme un projet européen, il se pourrait que ce soit la seule chance d'évaluer les programmes des pays qui n'ont jamais fait d'évaluation pour cause de manque de méthodologie.

3.4. Défis

Les experts se sont accordés sur le fait que le **procédé de recrutement** était un des plus grands défis dans les études d'évaluation multi-pays, particulièrement dans les cas où aucun financement additionnel n'était apporté. Ils ont souligné le fait qu'une telle étude ne peut être une réussite que si les chercheurs arrivaient à convaincre les professionnels, le personnel des programmes et autres acteurs-clés de sacrifier du **temps** et des **ressources** pour prendre part à la recherche. À leur avis, l'argument de combler les lacunes de connaissances dans les études d'évaluation des programmes d'aide aux auteurs de violences pourrait ne pas peser assez lourdement pour motiver les acteurs clés à participer. Ainsi, d'autres **bénéfices** devraient être mis en valeur comme des compensations pour leur participation aux études de recherche sur l'évaluation, tels que des formations externes pour le personnel des programmes (pas uniquement sur les méthodes d'évaluation), des visites d'études, des certificats (p.ex. délivrés par WWP-EN ou des organismes nationaux tels que Respect au Royaume-Uni ou BAG Täterarbeit HG, Allemagne). Selon les chercheurs parmi ces experts, un des bénéfices les plus importants d'une étude d'évaluation sera les commentaires reçus des participants qui pourraient alors amener à une amélioration de la qualité et de l'efficacité des programmes. En outre, les programmes participant à l'étude ont une plus grande chance de pouvoir bénéficier de fonds additionnels si un processus d'évaluation a été effectué et les résultats montrent que le programme a eu un impact. La dimension européenne d'une étude d'évaluation menée de manière externe augmenterait la crédibilité des résultats à un niveau national.

Un autre défi est de trouver **un accord** concernant l'usage de **modèles d'évaluation** identiques (critères, concepts, méthodes, outils). Puisque les programmes et les méthodes pratiques sont différents de pays à pays, des critères et des instruments différents sont, à ce jour, utilisés pour mesurer les résultats. Ainsi, le défi n'est pas seulement de sélectionner et utiliser les mêmes outils, mais de choisir des outils qui peuvent facilement s'adapter aux différents contextes.

3.5. Obstacles

En ce qui concerne les experts, effectuer une recherche de résultats multi-pays en Europe peut apporter un certain nombre d'obstacles. Un des obstacles majeur est le problème du **financement**, puisqu'une étude aussi complexe peut coûter très cher. Il est ainsi crucial de trouver un soutien financier pour un tel projet, mais cela peut s'avérer difficile. Un financement est nécessaire non seulement pour couvrir les travaux des équipes de chercheurs externes et autres coûts relatifs au projet (coûts de transport, équipement etc.) mais également pour couvrir les coûts des travaux supplémentaires effectués par le personnel des programmes, particulièrement dans les cas où les intervenants sont sous-payés ou sont bénévoles (ce qui est souvent le cas pour les programmes ayant lieu dans les anciens pays socialistes) et/ou dans les cas où l'évaluation des programmes n'est pas comprise dans le budget régional ou national.

Les **contextes différents** dans les différents pays européens est un autre obstacle qui influence autant le caractère des programmes que les procédés des futures recherches. La différence entre les contextes juridiques peut influencer le choix de méthodes spécifiques ainsi que les hypothèses de départ d'un projet de recherche. Les règles juridiques relatives au système d'intervention, la protection des données personnelles etc. peut empêcher les chercheurs **d'obtenir certaines informations** de certains contextes. Un exemple commun d'une telle situation a été donné par un expert espagnol, où les personnels des programmes d'aide aux auteurs de violences dans le système juridique n'ont pas accès aux victimes, et par conséquent, l'inclusion de la perspective des partenaires peut être difficile voire même impossible à réaliser.

Un autre aspect problématique est le **contexte culturel** des différents pays européens qui pourrait aussi influencer le processus de recherche. Dans certains pays, la **compréhension de termes cruciaux** tels que *violences conjugales, droits des victimes, comportement inapproprié, rôles des genres, égalité des genres* etc. peut différer des autres et ainsi, la conception d'un outil de recherche multi-pays peut s'avérer difficile. Par ailleurs, les pays se trouvent à des points différents en termes de situation **économique**, le rôle de la/des **religion(s)** (la proportion dans laquelle les idées religieuses influence la politique du pays, plus ou moins les pays séculaires), la variété des milieux culturels en termes de **migration** etc., mais aussi les degrés d'**intégration** des **politiques de genre** et d'une perspective de genre, le degré d'intégration et de développement des **programmes d'aides aux auteurs de violences**, ainsi que le niveau **d'acceptation de l'importance** d'un tel système d'intervention.

Pour les programmes eux-mêmes, les experts distinguent des différences importantes parmi les pays en termes de **théories du changement, critères de succès, types d'hommes** que l'on pourrait voir prendre part aux programmes d'aide aux auteurs de violences, et des approches différentes concernant le **genre des participants** dans les groupes. Il y a des programmes qui ne travaillent qu'avec des hommes, d'autres ayant une approche mixte ou qui séparent les hommes et les femmes en différents groupes, ou d'autre encore n'appliquant aucune différenciation. Pour résumer, selon les experts, l'obstacle majeur (méthodologique) à une étude de recherche d'évaluation multi-pays en Europe est que les **programmes d'aide aux auteurs de violences** sont en fait, pour une multitude de raisons, simplement très **différents**.

Enfin, certains experts doutent qu'il soit possible d'effectuer une étude européenne complexe puisqu'il n'y a toujours pas assez de programmes disponibles (dans certains pays comme la Hongrie, aucun programme n'est encore en place; Kutrovátz, 2013), ce qui pourrait mener à un problème d'échantillonnage sérieux.

3.6. Résultats sortis des discussions avec les experts et les groupes de contrôle

Les contributions des experts aux discussions de groupe et leurs réponses aux différentes questions ont été résumées en recommandations d'experts comme telles :

- En premier lieu, une **définition** de « programme d'aide aux auteurs de violences » doit être établie et reconnue dans les études d'évaluation multi-pays. Les programmes standards (coopération avec les services de soutien aux victimes et autres organisations, ceux basés sur une approche TCC ou psycho-éducative, perspective de genre) semblent les plus appropriés à ce moment précis pour des raisons également méthodologiques : nous avons besoin de programmes similaires dans des contextes différents, et non des programmes différents dans des contextes similaires. Les programmes doivent aussi pouvoir recueillir les données de résultats appropriées, ce qui dépend d'un niveau de coopération, c.-à-d. une approche coordonnée. De plus, l'accent de l'étude devrait être placé sur les programmes **compatibles** aux **standards** et **règlements** existants (p.ex. la Convention d'Istanbul).
- **Le projet de recherche**: de collaborer dans un nouveau projet financé semblait être le meilleur moyen d'aboutir à un projet d'évaluation multi-sites/multi-pays. L'idée initiale d'utiliser le guide pour l'évaluation en recueillant et analysant les données de façon centrale a été mise en doute, du moins à cette étape du processus. Néanmoins, le guide devrait être mise en place avec largesse, la recherche devrait se baser autant que possible sur ces outils, pour qu'il soit plus tard possible de revenir à l'idée initiale. Les différences du guide (produit résultant du Chantier 4 pour mesurer les résultats) et des instruments faisant partie de la méthodologie d'évaluation ont été soulignés par les experts. L'intégration des projets d'évaluation dans le travail quotidien semblait difficile.
- Les futures études devraient être conçues **de façon plus détaillées**, incluant le **système** utilisé lors de l'intervention. L'impact social des programmes d'aide aux auteurs de violences pourrait élargir le design initial concernant l'impact des programmes. Des **contextes** différents dans différents pays devraient être considérés depuis le départ, c.-à-d. la réglementation juridique, les aspects du système d'intervention et les aspects culturels.
- Les critères d'évaluation, la conception, les méthodes, les outils etc. devraient porter sur les **bonnes pratiques d'évaluation** actuelles et prendre en compte les avis des experts.
- Les nombreuses **questions** soulevées par les études d'évaluation devraient être traitées de **façon systématique**, préférablement en les liants à certains groupes de **parties prenantes**.
- Les modèles d'évaluation devraient montrer un fort degré de **transférabilité**, afin de pouvoir, à long terme, intégrer plusieurs projets venant de plusieurs pays.
- En ce qui concerne le type de programme : les moyens de référé (participation volontaire, envoyé par une organisation, ou une combinaison) peuvent être utilisés comme une variable et non un critère de sélection.
- Dans tous les cas, les **bénéfices** obtenus de la participation dans les études

d'évaluation devraient être mis en avant lors du processus de recrutement.

4. Comment avancer avec ces évaluations des PTAV ?

4.1. Point de départ : où en sommes-nous aujourd'hui ?

Après avoir étudié les différentes sources disponibles sur l'évaluation des PTAV et les points de vue des experts, il est devenu apparent que l'idée initiale du Chantier 3 doit être modifiée.

La planification, la conception et la performance, d'une manière exhaustive de larges études d'évaluation qui répondraient aux questions soulevées par les différentes parties prenantes se sont avérées plutôt difficiles, du moins pour le moment. En attendant, nous avons développé un concept d'efforts d'évaluation coordonnés, dont le but est d'offrir une série de projets d'évaluation plus petits, avec une méthodologie similaire et la possibilité de coordonner ces projets.

En résumé, les conditions à une évaluation multi-pays/projets multi site sont les suivantes :

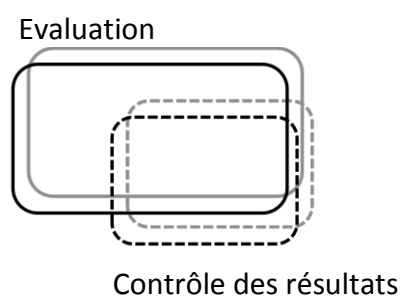
La **motivation principale** d'un tel projet est de profiter de la grande variabilité des conditions et des « systèmes » nationaux (pratiques des institutions, soutien des victimes, PTAV etc.) à travers l'Europe. Cette variabilité peut être une ressource pour les études d'évaluation, pour étudier les PTAV dans différents contextes. En outre, puisque beaucoup de programmes n'ont pas encore été évalués, une étude multi-sites/multi-pays peut être une opportunité d'élargir la base empirique relative aux PTAV. Comme nous l'avons démontré ci-dessus, il existe une grande diversité d'initiatives avec les auteurs de violences à travers les pays, mais on peut également trouver des programmes similaires (p.ex. TCC, PTAV basés sur les principes de Duluth), et des éléments de réponses communautaires coordonnés (RCC) sont répandus. Les procédures pour mesurer les résultats sont en place dans plusieurs pays, mais on trouve une grande variété méthodologique, qui rend les études actuelles difficiles à comparer. Il y a des différences et des problèmes relatifs à la terminologie et aux concepts (p.ex. « aboutissement », « succès » etc.) ainsi qu'au niveau de la conception (p.ex. en ce qui concerne les moments donnés pour prendre des mesures et les sources d'information), les groupes de contrôle sont difficiles à établir, sans parler des doutes éthiques concernant ces groupes de contrôle. Le taux de perte est un problème majeur en pratique, que ce soit les participants au programme ou les victimes-survivantes. Pourtant, comme l'ont conclu Hester et al. (2014) à partir de leur compilation d'études, « ... il serait possible, dans l'ensemble, de prendre des éléments des différentes approches afin de commencer à développer une méthodologie d'évaluation robuste. » (Hester et al., 2014, p. 39)

Les experts ont considéré qu'il était valable de poursuivre l'idée d'une approche multi-pays, et les différents groupes de parties prenantes sont désireux d'obtenir des réponses concernant l'efficacité des programmes pour différents types de PTAV. Comme l'affirment plusieurs parties prenantes, la priorité devrait porter sur les programmes qui sont compatibles aux standards et règles existantes (p.ex. la Convention d'Istanbul). Pourtant, il est apparu que les études d'évaluation qui seraient capables de répondre à ces questions demandent beaucoup d'effort et de ressources, et les administrateurs des projets ainsi que les professionnels se sont montrés sceptiques quant à l'intégration de ce type de travail dans leur routine quotidienne. En d'autres mots, il paraissait difficile de se lancer dans une étude d'évaluation internationale, multi-sites, multi-pays sans financement supplémentaire.

Les parties prenantes, en pratique, peuvent être convaincues de remplacer les outils qu'ils

utilisent pour évaluer les résultats de leurs travaux par des outils alternatifs, mais il reste un certain nombre de problèmes lors de l'évaluation du travail avec les auteurs de violences, et ces problèmes ne peuvent être résolus dans la pratique au jour-le-jour de ces programmes. Il est important de reconnaître que des études d'évaluation solides sont plus large dans leur concept et qu'elles requièrent plus de ressources que la mesure des résultats. Le travail et les efforts nécessaires dans les projets d'évaluation vont bien au-delà des routines quotidiennes et la mesure des résultats, bien que ces mesures peut être un bon point de départ et une fondation qui rendraient les PTAV « prêts à l'étude » (voir Gondolf, 2012; voir figure 2). Par conséquent, les praticiens mettent l'accent sur le besoin de fonds supplémentaires s'ils venaient à participer à des études d'évaluation.

Figure 2. Aspects communs et spécifiques des (a) études d'évaluation et (b) pratiques de contrôle des résultats.



En tout et pour tout, nous pouvons conclure qu'une base pour des projets d'évaluation multi-pays a été mise en place, mais une harmonisation de la méthodologie est définitivement nécessaire. Les PTAV et les services de soutien des victimes sont essentiels à l'évaluation des projets. Ainsi, pour augmenter les chances de succès, les études d'évaluation futures devraient être planifiées comme des projets séparés avec un budget suffisant pour couvrir les coûts des organisations prenant part.

4.2. La direction à prendre

Idéalement, dans quelques années, le problème du fractionnement dans les évaluations des PTAV aura été surmonté par l'harmonisation de la méthodologie d'évaluation ainsi les résultats des différentes études pourront être comparés et se compléter. L'idée est d'établir une institution coordonnée qui va promouvoir, soutenir et contrôler les projets d'évaluation des futurs PTAV autant dans la conception, les méthodes et les outils que dans la supervision du choix des questions spécifiques à être traitées. Cela ne sera pas un programme complet de recherche sur l'évaluation des PTAV, puisqu'il n'est pas réaliste d'anticiper et planifier de façon détaillé une série d'études qui durent plusieurs années, principalement à cause des restrictions budgétaires. Pourtant, il se peut fortement qu'une variété d'études d'évaluation soient effectuées, certaines à un niveau national, d'autres à un niveau international (p.ex. en tant que projets européens). Ces études-là doivent pouvoir être comparées et il devrait être possible de les imbriquer les unes dans les autres tel un « puzzle ».

Nous proposons d'utiliser l'organisation *Work-With-Perpetrators-European Network* (WWP-EN) comme institution coordinatrice, soutenue par le *Réseau Européen-WAVE* pour plusieurs

raisons : WWP-EN s'est placé en leader pour connecter les associations des PTAV en Europe et au-delà. Grâce à cette organisation centrale et la coopération des services de soutien aux victimes, des mesures dans l'établissement de standards pour un travail concret avec les auteurs de violences ont déjà été mises en place, et d'autres sont à venir. L'évaluation de ce travail n'est qu'une suite logique, tout comme l'est le contrôle des **standards** et comment effectuer cette **évaluation** (voir ci-dessous).

Cela correspond à ce que Gondolf appelle « les évaluations menées par les praticiens » (2012, p. 81), respectivement des projets d'évaluation qui sont soit initiés, soit conçus conjointement, soit influencés par ceux sur le terrain. Dans de nombreux cas, comme l'a souligné Gondolf, la coopération des chercheurs et praticiens s'est avérée difficile, dû en partie aux problèmes de langage, mais présente un fort potentiel de progrès.

4.3. Critères des projets d'évaluation

Comme nous l'avons appris des Chantiers 1 et 2, ainsi que des documents spécifiques, des standards méthodologiques et une terminologie commune semblent cruciaux à un développement concret de futurs projets d'évaluation. Evaluer le travail avec les auteurs de violences en général et les PTAV en particulier est une tâche très spécifique et les procédés tirés d'autres domaines (p.ex. l'évaluation en psychothérapie) ne peuvent être simplement « copiés-collés ». Certains des aspects propres à l'évaluation des PTAV sont :

- *Conception et terminologie*: les objectifs des PTAV ont été longuement discutés et dans ces discussions, des termes et concepts tels que « succès » ont été abordés. Différents groupes de parties prenantes ont exprimé leurs positions et leurs intérêts concernant les résultats des PTAV. Ces positions devront être prises en compte lors de la planification des études d'évaluations
- *Sources d'information*: les informations venant des hommes sont considérées comme insuffisantes par les experts dans le domaine (voir p.ex. Gondolf, 2012). Les informations données par les partenaires, la police ou les rapports de Justice sont nécessaires pour compléter les informations nécessaires. Les informations recueillies des partenaires sont considérées les plus fiables et les plus significatives et donc, devraient être les indicateurs centraux. Pourtant, cela a, à son tour, soulevé d'autres questions éthiques qui doivent être prises en considération (voir Downes, Kelly & Westmarland, 2014).
- *Moments d'observation*: Afin de parer au problème du taux de perte (des participants et des victimes-survivantes) d'un côté, et d'évaluer la durabilité des changements accomplis grâce aux programmes d'un autre, une simple pré/post conception n'est pas suffisante pour évaluer les PTAV. Hester et al. (2014) ont proposé un système de moments précis pour l'observation à être utilisé dans ce contexte. Ils ont défini quatre moments donnés : T0, lorsqu'un homme joint le programme ; T1, au début du programme ; T2, au cours de programme ; T3, à l'aboutissement du programme ; T4, au suivi, p.ex. 6 mois après la fin du programme. Ce système, accompagné des démarches à suivre pour recueillir des informations à chaque moment donné (qui doit être questionné, quels outils doivent être utilisés), peut constituer un critère central dans l'évaluation des PTAV. En plus d'une description détaillée du type d'hommes au moment T0 (p.ex. en ce qui concerne les caractéristiques sociodémographiques ou psychologiques) et par quel moyen ils se sont trouvés faire partie du programme, le problème du taux de perte pourra être étudié d'une bien

meilleure façon qu'il ne l'est aujourd'hui.

4.4. Promotion des critères

Une fois que de tels critères méthodologiques auront été établis, il faudra les promouvoir et les utiliser dans les études d'évaluation. Durant le Chantier 3, différents moyens de promotion ont été débattus. Les possibilités suivantes ont été identifiées :

- Un groupe de travail au sein du WWP-EN, responsable des problèmes d'évaluation, pourrait jouer un rôle clé dans l'harmonisation de la méthodologie de l'évaluation, le recensement des meilleures pratiques dans la recherche d'évaluation des PTAV et « l'orientation » des séries d'études d'évaluation. Les futurs projets auront besoin de coopérer avec d'autres programmes dans ce domaine, comme cela a été fait dans le passé. Ainsi, il est probable que les projets d'évaluation soient basés sur les organisations membre du WWP-EN, ce qui serait sensible sur le plan des critères d'évaluation du WWP-EN.
- De la même façon, les réseaux internationaux d'organisations de soutien aux victimes, en particulier le *Réseau Européen-WAVE*, joueront un rôle primordial. L'évaluation de PTAV demande la perspective des victimes, et ainsi, de tels projets nécessiteront le contact et la coopération non seulement des organisations membres du WWP-EN mais également celles membre de WAVE. Puisque WAVE et WWP-EN coopèrent, ces deux organisations européennes peuvent influencer grandement l'harmonisation de la méthodologie d'évaluation, via leurs organisations membres. De plus, il y a une possibilité de lancer des études d'évaluation (nationales et internationales) au sein du WWP-EN. En ce qui concerne les projets européens multi-pays en particulier, WWP-EN est l'organisation idéale pour coordonner les propositions dans le domaine de la recherche sur l'évaluation.
- L'étape suivante serait la promotion des standards méthodologiques à travers un projet d'évaluation financé par l'Union Européenne. La méthodologie concrète peut être développée dans le cadre de ce futur projet, sur la base de l'évaluation de la faisabilité du Chantier 3, une mise en commun de programmes de différent pays et d'apparence adéquats afin d'être identifiée par les données du Chantier 1 et le WWP-EN. Une fois que le projet aura pris la direction d'une évaluation multi-pays des PTAV, d'autres suivront et tireront parti de l'expérience du projet pilote.
- La promotion des pratiques basées sur les méthodes de recherche établies dans le guide du Chantier 4. Bien que ce guide se focalise plus sur le contrôle des résultats dans la pratique de tous les jours des programmes du WWP, elle représente un pas vers l'harmonisation de la collecte de données, dont des futures études vont pouvoir bénéficier dans un grand nombre de pays. Lorsqu'il y aura plus de programmes qui utiliseront ce guide, ils intégreront des routines basées sur les méthodes de recherche dans leur travail quotidien et cela préparera le terrain pour un projet d'évaluation. Le contrôle des résultats permettra aux équipes des programmes de « se préparer à la recherche ».
- Les programmes européens qui harmonisent la recherche et les pratiques à travers l'Union Européenne, offrent aussi des possibilités de méthodes d'harmonisation dans la recherche sur l'évaluation. Il existe des programmes européens qui visent à développer un Espace Européen de la Recherche, p.ex. COST. Au niveau de la pratique, les réseaux WWP-EN et WAVE joueront un rôle clé dans la promotion des standards méthodologiques dans l'évaluation, quand ils aideront à choisir les

programmes et encourageront leurs membres à participer à une étude d'évaluation multi-pays.

4.5. Une approche exhaustive vers l'évaluation des PTAV et le rôle des études multi-pays

En partant de la métaphore du « puzzle » mentionné plus haut, différents types d'études d'évaluation peuvent répondre à différentes questions. L'important est de s'assurer que ces études peuvent être comparées entre elles et mises en relation. Cela peut se faire en utilisant une méthodologie commune (ou du moins un tronc commun de procédures et d'outils comparables). Nous proposons d'observer les différents types d'études d'évaluation avec la perspective de la catégorisation, illustrée ci-dessous (des informations plus détaillées sur ce point et autres aspects seront fournies dans des articles séparés rédigés et publiés dans le cadre du projet IMPACT).

4.5.1. Études basiques à contexte unique

Ces études-là évaluent l'impact d'un PTAV dans le cadre d'un RCC. De telles études sont effectuées sur le terrain, et non en conditions contrôlées, et par conséquent, il est probable qu'elles appliquent un concept quasi-expérimental car l'affectation aléatoire des unités de recherche (p.ex. les hommes) aux conditions de traitement n'est souvent pas possible, pour des raisons méthodologiques et éthiques. Les abandons ont été utilisés comme groupe de contrôle, ou encore d'autres groupes d'hommes qui n'ont pas complété les interventions (p.ex. ceux étant été exclus des programmes ou n'y ayant pas été admis) et ont été comparés à ceux qui ont terminé leur programme.

Nous voulons, ici, proposer trois éléments qui, dans le cadre d'une future méthodologie d'évaluation, auront besoin d'être améliorés, par *les équipes de travail du WWP-EN* par exemple : (a) observations/moments donnés; (b) conception; (c) indicateurs de succès.

(a) Observations / moments donnés

Hester et al. (2014) ont proposé, dans leurs analyses des études d'évaluation européennes, un système pour décrire « la forme et la taille des échantillons à différents moments du processus d'évaluation » (p.17). Ce système devrait être utilisé dans les études d'évaluation des PTAV afin de définir les différents moments pour l'observation. Le système consiste en cinq moments, accompagnés de propositions de sources d'information à chaque moment précis respectivement.

- T0: « pré-départ ». Il est important de connaître le types d'hommes qui sont envoyés à ces programmes (données sociodémographiques, caractéristiques psychologiques, abus de substances, moyen de référence etc.) et si oui ou non ils rejoignent les programmes. La comparaison des échantillons aux moments T0 et T1, avec les raisons des abandons, peut fournir des informations intéressantes pour le développement d'interventions pour ces hommes qui ne rejoignent pas de programme. Un grand nombre de ces informations peuvent être obtenues des victimes-survivantes, certaines des hommes, d'autres des professionnels (pour en savoir plus sur les aspects éthiques de l'évaluation des programmes par les victimes-survivantes, voir Downes, Kelly & Westmarland, 2014.). En ce qui concerne les **victimes-survivantes**, les indicateurs d'une intervention réussie (voir ci-dessous) doivent être relevés au moment T0. Ces indicateurs seront les mesures de réussite de référence aux moments

T3 et T4.

- T1: « Départ ». la comparaison des échantillons aux moments T0 et T1 informera sur la première vague d'abandons, la comparaison aux moments T1, T2 et T3 respectivement montrera quels hommes continuent le programme et lesquels l'ont quitté. Les changements dans la vie des hommes et les informations sur les variables contextuelles seront importantes des moments T0 à T3, de façon à pouvoir interpréter les différents abandons. En ce qui concerne les **hommes**, l'accent, à ce moment, sera mis sur les variables qui devraient changer grâce au programme (mesure des résultats pour le traitement), observées au moment T1 (p.ex. comportement, attitudes, motivation, facteurs de risque etc.). Là encore, on peut questionner les **victimes-survivantes** pour identifier les indicateurs d'une intervention réussie.
- T2: « Pendant ». il est conseillé de prendre des mesures régulièrement durant l'intervention. Le développement à long-terme des indicateurs peut apporter des informations importantes concernant aussi bien les hommes que les victimes-survivantes. De plus, les questions relatives aux procédés peuvent être adressées et liées aux résultats.
- T3: « fin ». La comparaison des groupes de traitement et des groupes de contrôle aux moments T1 (avant le traitement) et T3 (après le traitement) est le noyau central des études d'évaluation de l'impact des interventions. À nouveau, l'analyse devrait incorporer au moins deux perspectives : celle des victimes-survivantes, pour identifier les indicateurs de succès, et celle des hommes participant aux programmes, respectivement le groupe de contrôle, pour mesurer les résultats des traitements mêmes. Les informations des professionnels et, par addition, les rapports officiels peuvent être recueillis au moment T3.
- T4: « Suivi ». la stabilité des changements est évaluée lors d'une période de suivi, idéalement 6 mois après l'aboutissement du programme. Là encore, les victimes-survivantes et les hommes ayant participé aux programmes devraient être interviewées, les professionnels et rapports officiels pouvant constituer des sources d'information additionnelles.

Une analyse des abandons peut être effectuée à chacun des moments définis, pour les hommes ainsi que pour les victimes-survivantes, à condition que les informations pertinentes (p.ex. les variables sociodémographiques, les risques, les besoins etc.) aient été relevées au moment T0.

(b) Conceptions

Les Études Randomisées Contrôlées (ERC) sont perçues comme la norme de référence dans la recherche clinique et psychologique. Les ERC dépendent de l'affectation randomisée des unités de recherche (p.ex. les hommes) aux conditions de traitement (p.ex. l'intervention contre la non-intervention). Il y a beaucoup de problèmes liés à cette procédure (p.ex. des problèmes éthiques), et les ERC effectuées ont connu un grand nombre de critiques spécifiques (voir Gondolf, 2002, 2012). Même dans le système judiciaire, il est souvent difficile d'effectuer des ERC (voir p.ex. Dobash et al., 2000), même si les meilleures conditions pour effectuer une ERC est probablement en milieu carcéral (voir ci-dessous).

Figure 3. Conception d'une Étude Randomisée Contrôlée (ERC)

Affectation	T1 (pré)	(T2, pendant TM)	T3 (post)	(T4, suivi)	
randomisée	Groupe de traitement	O	X	O	O
des individus					
à un groupe	Groupe de contrôle	O	O	O	O

Par conséquent, dans plusieurs études, les abandons, les hommes qui n'assistent pas aux sessions où ceux qui ont été exclus des programmes ont été définis comme des groupes de contrôle, tandis que ceux qui ont complété les PTAV représentent les groupes de traitement. Ainsi, les groupes de traitement et les groupes de contrôle ne sont pas équivalents et cela constitue une menace à la validité interne de telles études (Shadish, Cook & Campbell, 2002). Dans sa fameuse étude d'évaluation, Gondolf (2002) a remarqué des impacts mineurs pour les hommes qui ont complété les programmes, comparé aux abandons. Afin de parer au problème de cette non-équivalence, Gondolf a utilisé l'analyse des coefficients de propension et l'analyse des variables instrumentales pour ajuster les résultats. Ces procédés ont permis d'obtenir une plus grande efficacité. Le principe des tels procédés d'ajustement est comme suit : la différence entre les résultats des groupes de traitement et des groupes de contrôle non-équivalents est corrigée statistiquement en utilisant la différence entre les groupes (qui doit être connue au début de l'étude) comme une sorte de variable de contrôle. Par conséquent, on obtient une valeur de l'effet qui reflète les effets que l'on peut espérer, en supposant que les deux groupes étaient aussi similaires que possible. Cette valeur de l'effet corrigée est considérée, par les experts en statistiques, comme plus appropriée que la valeur initiale, non corrigée (et la valeur de l'effet non corrigée et considérée comme inappropriée).

Beaucoup d'études d'évaluation antérieures ont mis l'accent sur les hommes participant aux programmes. Dans les études où les victimes-survivantes étaient incluses comme source d'information concernant l'impact du programme, les chercheurs ont souvent pris contact avec les auteurs de violences en premier lieu, et ont tenté par la suite de contacter les victimes-survivantes respectives. Cette approche peut aussi être suivie dans le futur. Comme l'a démontré Gondolf (2002) dans son étude, jusqu'à 60% des victimes-survivantes initiales ont pu être contactées au moment du suivi.

Pourtant, il existe des approches alternatives, comme celles qui suivent qui ont été inspirées par le projet MIRABAL. L'emphase de la recherche peut être déplacée des participants au programme vers les victimes-survivantes dans les centres de soutien aux victimes. Étant donné qu'il y a des régions géographiques où l'on peut trouver des centres de soutien aux victimes, mais avec ou sans PTAV, l'effet supplémentaire d'un PTAV peut être étudié quand les résultats des victimes-survivantes dans de telles régions ou villes sont comparées à ceux d'une région/ville sans PTAV (les centres de soutien aux victimes sont existants dans les deux types de régions). Dans un tel concept, les victimes-survivantes bénéficient d'un soutien et d'un suivi à long terme (T1 à T4), tout comme leur partenaire qui sont référés à un PTAV lorsque ce dernier est disponible (groupe de traitement), ou qui ne peuvent être référés à un PTAV car il n'y en a pas (groupe de contrôle). En se basant sur les informations données par les victimes-survivantes concernant leur partenaire, il serait même possible d'effectuer une sorte de regroupement des groupes de traitement et des groupes de contrôle au moment T0.

Bien qu'il paraisse utile et sensé d'inclure la perspective des hommes dans l'évaluation et de continuer en suivant le système ci-dessus (observations/moments donnés), nous pourrions même considérer ne prendre que les perspectives des victimes-survivantes en compte, avec

⁴ Les résultats du projet MIRABAL sera présenté début 2015. La méthodologie et l'approche du projet devraient être examinés par les futurs groupes de travail d'évaluation du WWP-EN.
<https://www.dur.ac.uk/criva/projectmirabal/>

quelques informations additionnelles obtenues des professionnels concernant les abandons, « les doses » ou l'aboutissement des programmes. Dans ce cas, le PTAV est perçu comme une « boîte noire », administrée à un groupe d'auteurs de violence à une certaine dose (en fonction de la durée de leur participation). Le groupe de traitement (les victimes-survivantes dont les partenaires participent à un programme) est alors comparé au groupe de contrôle (les victimes-survivantes dont les partenaires ne participent pas à un programme). Une condition préalable demandera que les contextes des deux régions soient très similaires, pour que les différences entre les groupes de traitement et les groupes de contrôle puissent être attribués à l'existence ou l'absence d'un PTAV. Bien que les procédés détaillés soient plus complexes⁵, l'idée de base d'étudier l'impact d'un PTAV par les résultats différentiels des victimes-survivantes qui font partie d'un programme de soutien aux victimes dans différentes régions (avec ou sans PTAV) est très prometteuse.

De façon similaire, d'autres concepts peuvent également être développés, p.ex. des modèles chronologiques interrompus, où les victimes-survivantes sont suivies sur une certaine période de temps, relativement longue, qui correspond à T0 - T5, et où, à un certain moment, un PTAV est mis en place dans la région. L'hypothèse est que les résultats qui sont relatifs à la situation de vie des victimes-survivantes s'amélioreront une fois que leur partenaire masculin est référé à un PTAV de façon systématique. Les modèles chronologiques sont aussi valables pour les recherches à cas unique. Cette possibilité ne devrait pas être sous-estimée, particulièrement dans les contextes où les échantillons sont très petits. La méthodologie du cas unique est un développement prometteur également au niveau des statistiques et peut être utilisée dans le contexte d'une évaluation des PTAV. Il est tentant de considérer comment les effets des PTAV peuvent être étudiés en se basant sur les données d'un cas unique. Cette approche est particulièrement utile lorsque différentes personnes reçoivent des traitements différents et à différentes intensités (« doses »), en fonction de leurs besoins individuels.

Le projet MIRABAL a annoncé qu'il allait apporter une nouvelle approche quantitative progressive à l'évaluation des programmes d'aide aux auteurs de violences, sous le nom d'*analyse de survie*. Cette forme d'analyse provient de la recherche médicale et du calcul de l'espérance de vie, mais elle est également utilisée dans d'autres domaines. L'analyse de survie commence avec des tables de mortalité d'où l'on calcule les probabilités cumulatives de survie. Ces probabilités, généralement illustrées par des graphiques sur une certaine période de temps, peuvent être adaptés pour différents sous-groupes, p.ex. les personnes recevant des traitements différents. De plus, les effets des co-variables peuvent y être observés. Nous conseillons l'adoption de cette approche innovante par les futurs groupes de travail d'évaluation du WWP-EN.

Les résultats des études d'évaluation de base, comme elles le sont décrites ici, feront office de déclarations concernant les effets additionnels d'un PTAV dans une RCC (comparé à une « RCC sans PTAV»). Quel que soit le modèle suivi, des études différentes ne peuvent être comparées que si le système T0-T4 est utilisé de façon uniforme (connaître le type d'hommes participant aux différents programmes, qui abandonne et quand), et si les mêmes mesures de résultat sont utilisées.

⁵ Par exemple, les groupes doivent être définis ultérieurement : les abandons du côté de victimes-survivantes tout comme les participants ; les hommes qui finissent le programmes et leur partenaires/ex ; les hommes exclus et ceux dont on a refusé l'admission etc.

⁶

(c) Critères de succès et mesure des résultats

Dans un projet de recherche précédent pour le projet MIRABAL, Westmarland, Kelly et Chalder- Mills (2010) ont proposé les « critères de succès » des PTAV suivants:

- Amélioration des rapports entre les participants aux programmes et leur partenaire/ex-partenaire ; respect et meilleure communication.
- Élargissement des possibilités pour les partenaires/ex-partenaires d'agir, de s'exprimer, de prendre des décisions ; amélioration de leur qualité de vie.
- Amélioration de la sûreté des femmes et des enfants, libérés de la violence et des abus.
- Comportement parentale positif, sûr et partagé.
- Pour les hommes, prise de conscience de soi et des autres, compréhension de l'impact de la violence sur les partenaires et les enfants.
- Enfance plus sûre et saine. Les enfants se sentent entendus et valorisés.

Ces critères étaient utilisés pour formuler des questions et des problèmes concrets pour l'étude MIRABAL (voir Westmarland et Kelly, 2012). Dans cette approche, la perspective des victimes-survivantes et des enfants est la source d'information principale pour évaluer l'impact d'un PTAV. Idéalement, cette information est complétée par la perspective des hommes participant aux programmes, avec les critères de succès mentionnés ci-dessus. Les professionnels et les rapports officiels qui fournissent des informations supplémentaires, peuvent être utilisés pour répondre aux questions qui vont au-delà de l'évaluation des résultats des PTAV. Pour fixer les futures études d'évaluation dans un exemple de bonne pratique, nous conseillons aux groupes de travail d'évaluation du WWP-EN de prendre ces indicateurs en considération.

4.5.2. Études multi-sites à contextes multiples

Dans les *études basiques à contexte unique*, l'objet était de déterminer l'impact des PTAV, lorsque les conditions sont plus ou moins les mêmes dans deux localités. Nous voudrions maintenant nous tourner vers un autre cas, lorsque des PTAV *similaires* opèrent dans des *conditions différentes*. Un schéma de représentation de cette approche est illustré dans la figure 4. Le contexte d'un PTAV est constitué d'éléments locaux pertinents qui font partie du système d'intervention en cas de violences domestiques (le police, les tribunaux, les services de protection des victimes etc.), ainsi que des conditions à un niveau plus élevé, p.ex. le problème de financement au niveau régional, ou encore la législation, les campagnes de sensibilisation et les mesures préventives à un niveau national. Par ailleurs, des conditions internationales et globales peuvent influencer les résultats du système.

Dans les *études basiques à contexte unique*, nous avons supposé que le contexte est une constante, et qu'il y a deux localités (avec et sans PTAV) dans un pays donné, avec des contextes locaux et régionaux identiques ou très similaires. Dans les études multi-sites à contextes multiples, nous supposons que des PTAV très similaires sont mis en œuvre dans des conditions différentes, comme on peut s'y attendre de différents pays. Les résultats du

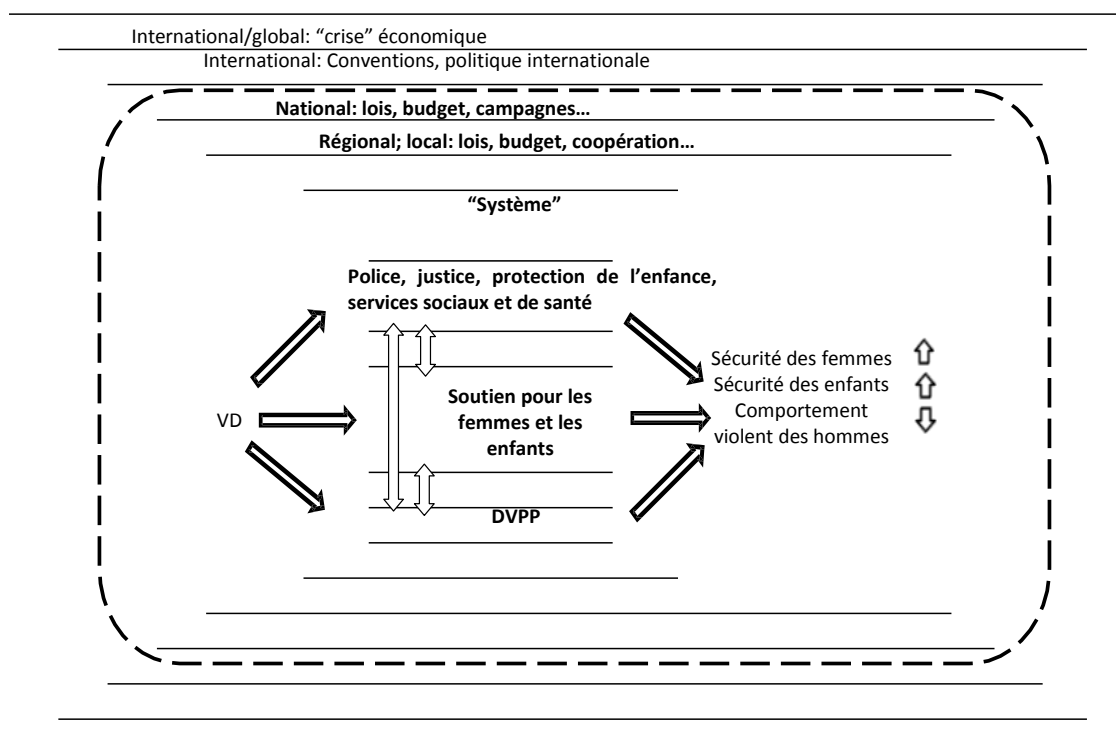
Chantier 1 ont montré que des programmes similaires existent dans différents pays, probablement dû au fait que beaucoup de programmes partagent les mêmes racines, telles que le modèle Duluth ou le programme CHANGE. Quand on peut identifier des PTAV similaires dans deux pays ou plus, et de préférence ayant les mêmes systèmes de référés (ordonnance du tribunal, participation volontaire ou les deux), les résultats des interventions des systèmes tous entiers peuvent être comparés. Dans ce cas, nous pouvons supposer que la façon dont les PTAV fonctionnent est constante, mais que le contexte est différent, menant à des résultats différents. Dans un tel concept, le *rôle du contexte* et son *impact sur les résultats d'un PTAV* sont comparés et étudiés. C'est dans cette situation exacte que l'approche multi-pays prend du sens.

Encore, afin de pouvoir lier une telle étude à d'autres évaluations, il est nécessaire d'appliquer la même méthodologie centrale à toutes les études. Après avoir effectué une étude multi-sites à contexte multiple, nous pouvons nous attendre à des déclarations concernant l'influence du contexte sur les résultats d'un certain genre de PTAV.

Si le contexte doit être analysé plus en profondeur, et ne pas seulement être considéré comme une « boîte noire » (avec un impact général connu, mais un impact inconnu sur les éléments individuels), le défi sera de *mettre le contexte en œuvre* afin de déterminer l'influence des éléments sur les résultats des PTAV. En prenant la figure 4 comme point de départ, la tâche est de transformer le contexte en variables pour une analyse quantitative. Différents composantes doivent être « traduites » d'une façon qui leur permettra de s'intégrer aux procédures de statistiques quantitatives :

- Au niveau local : soutien pour les femmes et les enfants ; le travail de la police, du système juridique, des systèmes de protection de l'enfance ; services sociaux et services de santé ;
- Au niveau régional et national : moyens juridiques, moyens de prévention de la violence, situation budgétaire ; situation générale de l'égalité des genres.

Figure 4. Schéma pour différents types d'études d'évaluation des PTAV.



Note. VD... Cas de violences domestiques; PTAV... Programme de d'aide aux auteurs de violences. Les cas de violences domestiques rencontrent un système d'interventions qui consiste de différents éléments en interaction. Les résultats des interventions du système dépendent de chaque élément et de leurs interactions, autant à un niveau régional que national ou même international. En prenant le point de vue d'un PTAV, les autres éléments du système et les conditions des niveaux supérieurs (nationaux, internationaux) constituent le contexte des PTAV.

Une telle mise en œuvre est encore absente aujourd'hui, et son développement nécessiterait un grand nombre de ressources. Ici, nous ne voulons que donner des indications sur comment cette mise en œuvre pourrait être faite, comme une préparation pour de futures *études multi-sites à contextes multiples*.

Au *niveau du système*, les réactions des tribunaux ont été identifiées comme des influences pertinentes sur les résultats des PTAV, dans les cas où les participants sont sous ordonnance du tribunal (p.ex. Gondolf, 2002). Pour mettre cette influence en œuvre, on peut commencer par les perspectives des participants au programme, des professionnels, des victimes-survivantes ou des experts. Si on se base sur la perspective des *hommes participant au programme sous ordonnance juridique, évalués par les professionnels*, les indicateurs suivants peuvent être utilisés :

A. Les expériences des participants et leurs attentes sur les réactions de la Cour.

- « en cas de non-respect, il y a eu (ou on peut s'attendre à) une réponse rapide et conséquente de la Cour » (Value=1)
- « en cas de non-respect, il y a eu (ou on peut s'attendre à) une réponse lente ou irrégulière de la Cour » (0,5)⁶
- « en cas de non-respect, il n'y a eu aucune réponse de la cour (ou il n'y en aura aucune). » (0)

B. Signification subjective des conséquences annoncées ou imposées.

- « les conséquences par les tribunaux sont significatives / sévères aux yeux des participants. » (1)
- « les conséquences par les tribunaux ont une certaine négative signification aux yeux des participants. » (0,5)
- « les conséquences des tribunaux n'ont aucune signification pour les participants. » (0)

Les informations au cas par cas concernant A, B, A*B est nécessaire pour tous les éléments d'un système et par différents points de vue (les hommes, les professionnels, les victimes-survivantes). De cette manière, un grand nombre d'informations relatives à chaque cas peut être enregistré dans une matrice de données utilisée pour effectuer p.ex. une analyse multi-niveaux pour évaluer l'influence des variables et des niveaux sur les résultats du PTAV. À un niveau supérieur, les experts peuvent contrôler les réponses des tribunaux *en général*, information qui est, à nouveau, insérée sous forme de variable dans la matrice de données. Dans une telle étude, les influences appropriées seront déterminées de façon empirique.

Pour déterminer quels sont ces éléments, on peut commencer à développer un outil de comparaison des contextes en partant des recherches comparatives des mesures juridiques et de soutien, telles que l'EIGE (2012). Il existe également des projets antérieurs qui avaient pour but de situer et comparer les contextes concernant les mesures pour les victimes-

survivantes dans les différents pays européens (p.ex. EUCPN, 2013). Ce travail montre quelles mesures juridiques et pratiques sont en place à travers l'Europe.

⁷ Ou un autre plage d'alternatives entre 0 et 1.

Un autre exemple d'un indicateur potentiellement utile est l'Index d'Égalité de Genre (GIA – Gender Impact Assessment), développé par l'EIGE (2012). Bien que la violence ne fasse pas partie du GIA en raison d'un manque de données comparables, le GIA peut être utilisé en tant qu'index général pour déterminer la position d'un pays concernant l'égalité de genre, et ainsi l'entrer dans l'équation en tant que variable de contrôle ou indépendante.

Une autre approche intéressante a été proposée par Hester et Lilley (2014), qui mettait en évidence la nécessité de perturber les différents facteurs de risque de violences à différents niveaux au sein d'un cadre écologique (société, institutions, famille/pairs, individus). Cette proposition est basée sur un modèle interactif de facteurs de risques développé par Hagemann-White et leurs collègues (voir la Commission Européenne, 2010). Leur modèle peut être utilisé pour définir les démarches à prendre dans la prévention des violences entre partenaires. Les éléments préventifs peuvent alors être mis en œuvre et utilisés comme variables dans les analyses multi-niveaux.

4.5.3. Études multi-sites à contexte unique

Le but de ce genre d'études est de comparer différents types de PTAV au sein d'un seul contexte. Il est difficile de déterminer ce qui mène à différents résultats quand on étudie *différents* programmes dans *différents* contextes. Il est donc conseillé de garder le contexte comme une constante, p.ex. en comparant deux types de programmes dans une région.

Cette approche n'adresse pas seulement les différentes façons de travailler avec les auteurs de violences (p.ex. plus penchés sur l'approche TCC ou psychodynamique, avec ou sans des services supplémentaires tels que le traitement de l'alcoolisme ou la psychothérapie), mais elle permet aussi de comparer différents moyens de référés, p.ex. les programmes TCC dans une région où un programme travaille avec des participants sous ordonnance juridique et un autre avec des participants qui ont approché le programme volontairement etc. Le but ne devrait pas être de simplement déterminer quel programme est le plus efficace, mais quelle approche fonctionne le mieux avec quel type de participants, ce qui signifie qu'il est nécessaire d'étudier les caractéristiques des hommes au moment de l'admission. Le débat sur la question de donner un traitement différent à différents types d'auteurs de violences peut être traité lors d'études qui comparent différents programmes (au sein d'un seul contexte), et répondra à la question de l'intensité du traitement. Les programmes standards peuvent fournir un point de référence contre lequel d'autres programmes contenant de nouveaux éléments peuvent être comparés.

Là encore, il est essentiel d'utiliser la même méthodologie au cours de toutes ces études, et dans cet exemple, les informations recueillies au moment T0 concernant les hommes participant au programme seront d'une pertinence cruciale. Les études multi-sites à contexte unique sont l'approche correcte pour développer notre connaissance de l'interaction des paramètres, des méthodes, des types d'auteurs de violences etc. Par conséquent, nous pouvons attendre des conclusions concernant les effets différentiels de différents types de PTAV pour différents types d'auteurs de violences.

4.5.4. Études en conditions de pseudo-laboratoire

Finalement, il y a un type d'études d'évaluation dans lequel il est inutile et parfois impossible (p.ex. légalement) d'incorporer les perspectives des victimes-survivantes (p.ex. dans le cas des programmes d'aide aux auteurs de violences en milieu carcéral en Espagne). D'un autre côté, durant les recherches pour le Chantier 2 (Hester et al., 2014) ; les seuls cas d'ERC ont été trouvés dans des institutions fermées telles que les prisons ou les hôpitaux psychiatriques. Les prisons empêchent l'observation immédiate de l'impact des programmes d'aide aux auteurs de violences car les détenus ne vivent pas avec leur partenaire et ont un contact limité avec eux en général. Néanmoins, dans le cadre d'un programme de réhabilitation, les PTAV effectués en prison préparent les participants à une vie après l'emprisonnement, et doivent être considérés comme des interventions significatives et importantes.

En termes d'évaluation des PTAV, le rôle des programmes en milieu carcéral peut être comparé à une situation en laboratoire pour des expériences psychologiques. Dans les prisons, les ERC sont en principe possibles. Les concepts de listes d'attente en particulier (c.-à-d. les personnes qui ne reçoivent pas de traitement et qui font partie d'un groupe de contrôle bénéficieront du traitement plus tard) sont une manière viable d'effectuer des ERC dans les prisons sans implications éthiques majeures. Ainsi, en se concentrant sur les différents éléments des traitements et en les étudiant de manière systématique, les études en milieu carcéral peuvent aider à améliorer les programmes. Des éléments efficaces peuvent être mis en place dans des programmes dans ce domaine (et peuvent aussi être testés et validés par les études multi-sites à contexte unique).

Au vu des résultats des études en conditions de pseudo-laboratoire, on peut s'attendre à obtenir des énoncés différenciés concernant les effets des traitements dans un milieu carcéral (basés principalement sur les perspectives des hommes), ainsi que sur l'efficacité de certains de ces éléments de traitement. On peut effectuer des études de corrélation avec les cas de récidive, bien que la récidive dépende de plusieurs facteurs et influences, et non des résultats d'apprentissage d'un homme ayant participé à un programme en prison, et donc les attentes concernant ces types de validation ne doivent pas être trop élevés.

5. Conclusion et prochaines étapes

L'idée d'une grande étude d'évaluation multi-pays qui profiterait de la diversité des approches du travail avec les auteurs de violences en Europe, et qui répondrait à une gamme de questions d'un coup, a dû être modifiée. Nous nous sommes concentrés sur le travail avec les auteurs de violences conjugales et sur des approches coordonnées, pour des raisons variées, mais aussi car des experts venant de la recherche et de la pratique nous ont conseillé de suivre cette route lors des ateliers organisés pour le projet. Nous avons souligné les bases d'une méthodologie pour ce domaine, qui nécessite encore un développement plus détaillé, afin de pouvoir avancer vers des procédures d'évaluation des PTAV plus systématiques et harmonisées, tant à un niveau national qu'international. La nouvelle organisation internationale WWP-EN peut et devrait jouer un rôle important car elle a la possibilité de développer et de promouvoir les standards d'une évaluation des PTAV parmi ses membres, et beaucoup de futurs projets d'évaluation auront besoin de la coopération de ces derniers. Une étape hautement importante qu'il faut prendre aujourd'hui est de désigner un groupe de chercheurs et de praticiens intéressés au sein du

WWP-EN qui se sentiraient responsable pour cette tâche. Lorsqu'ils développeront les méthodes et les standards, ils pourront se baser sur les résultats du projet IMPACT en général, et les recommandations du Chantier 3 en particulier.

Il y a certaines spécificités concernant l'évaluation de PTAV. Une des plus importantes est l'idée que les informations les plus fiables concernant les changements proviennent des partenaires et des enfants des auteurs de violences. Par conséquent, la coopération avec les organisations de protection des victimes est nécessaire, au mais aussi au niveau d'un réseau avec des organisations cadres telles que WAVE. Il s'est trouvé qu'une approche systémique est la plus appropriée, ce qui signifie que nous ne devons pas nous efforcer d'évaluer des programmes d'aide aux auteurs de violences isolés, mais d'adresser le système d'intervention tout entier. Les manières de procéder à cet égard ont été exposées dans ce document de travail, adoptant de nouveaux développements et prenant exemple des bonnes pratiques dans l'évaluation de PTAV. Nous avons proposé un cadre de travail pour des études d'évaluation qui permettent de lier et comparer des études uniques et d'apporter différents concepts à un système sensé, basé sur les développements des autres Chantiers du projet IMPACT, des études antérieures et certains développements prometteurs dans l'évaluation des PTAV. Nous espérons que cette proposition contribuera à surmonter une situation diffuse et plutôt problématique en ce qui concerne la comparabilité des évaluations de PTAV aujourd'hui. Petit à petit, nous devrions arriver à détecter et à démontrer l'ampleur des effets des interventions optimisées (« qui a besoin de quoi ») autant pour les auteurs de violences que pour les victimes-survivantes. Pour obtenir une valeur fiable de l'ampleur des effets des PTAV, il faut également considérer les bases des analyses coûts/bénéfices et coûts/efficacité significatives. Dans les secteurs sociaux et médicaux, de telles analyses prouvent normalement que les ressources allouées à l'intervention ne sont pas uniquement une question d'éthique, mais elles permettent également d'économiser les ressources publiques. Il est commun de trouver que les coûts des interventions sont faibles comparés aux coûts engendrés par l'inaction.

Davantage d'aspects et de considérations seront publiés dans des documents séparés, basés sur le travail effectué au cours du Chantier 3 du projet IMPACT.

6. Bibliographie

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.
- Conseil de l'Europe (2011). *Convention du Conseil de l'Europe sur la prévention et la lutte contre la violence à l'égard des femmes et la violence domestique*. Extrait de <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. Rapport*. Vilnius: European Institute for Gender Equality. Extrait de <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: Bruxelles.
- Commission Européenne (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: Office des Publications de l'Union Européenne.
- Geldschläger, H., Ginés, O., Nax, D. & Ponce, A. (2014). *Outcome measurement in European perpetrator programmes: A survey*. Extrait de 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: Conseil de l'Europe.
- Hester, M., Lilley, S.-J., O'Prey, L. & Budde, J. (2014). *Overview and analysis of research studies evaluating European perpetrator programmes*. Extrait de

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*. Extrait de 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*. Londres et Durham: London Metropolitan University et Université de Durham.
- 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?* Londres: Respect.
- Organisation Mondiale de la Santé (2002). *Rapport mondial sur la violence et la santé. Résumé*. Extrait de 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*. Extrait de <http://www.work-with-perpetrators.eu/index.php?id=86>