



Possibilities for multi-site/multi- country European evaluation studies on domestic violence perpetrator programmes

Working paper 3 from the Daphne III project “IMPACT: Evaluation of European Perpetrator Programmes”

Christian Scambor, Katarzyna Wojnicka & Elli Scambor



Index

1. Introduction.....	3
2. Methods	5
2.1 Analysis of literature	5
2.2. Applying the results of Workstream 1.....	6
2.2.1. Important aspects from Workstream 1.....	7
2.2.2. Conclusion	9
2.3. Applying the results of Workstream 2.....	10
2.3.1. Important aspects from Workstream 2.....	10
2.3.2. Conclusion	12
2.4. Experts' focus groups interviews.....	12
3. Results: Experts' perspectives	14
3.1. Questions.....	14
3.2. Possibilities	15
3.3. Advantages	16
3.4. Challenges	16
3.5. Obstacles	17
3.6. What results from the expert focus groups and the discussion	18
4. How to move on with the evaluation of DVPPs	19
4.1. Starting point: Where we are at the moment.....	19
4.2. Where we should head towards	20
4.3. Standards for evaluation projects	21
4.4. Promotion of standards.....	22
4.5. A comprehensive approach towards evaluation of DVPPs and the role of multi-country-studies	23
4.5.1. Basic single-context studies.....	23
4.5.2. Multi-site multi-context studies	27
4.5.3. Multi-site single-context studies	30
4.5.4. Laboratory-like studies	30
5. Summary and next steps	31
6. Literature	32

List of frequent abbreviations

CBT ... Cognitive-behavioural therapy

CCR ... Co-ordinated community response

DVPP ... Domestic violence perpetrator programme

RCT ... Randomized Controlled Trial

T0 ... Point of time for observation: At referral; intake-phase

T1 ... Point of time for observation: "Pre", at the start of the intervention (e.g. group work)

T2 ... Point of time for observation: "During", during the intervention (e.g. group work)

T3 ... Point of time for observation: "Post", at the end of the intervention (e.g. group work)

T4 ... Point of time for observation: "Follow-up", a period of time after the end of the intervention (e.g. 6 months after the end of the group work)

IMPACT ... Daphne III project "Evaluation of European Perpetrator Programmes"

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

WWP ... Work with perpetrators

WWP – EN ... Work With Perpetrators – European Network (international organisation)

1. Introduction

According to the experts in the field, evaluation of perpetrator programmes is highly important for practitioners and the programmes' staff in order to monitor and improve the quality of their work, as well as for policy makers to help making strategic decisions (Geldschlager et al., 2014; Hester, Lilley, O'Prey, Budde, 2014). Unfortunately, to this day, only a few evaluation projects on European programmes have been published in scientific journals – even though in their review in Workstream 2, Hester et al. (2014) have found at least 65 national studies (published in journals or as “grey literature”). However, the studies appear fragmentary. The majority of projects were carried out at one site and only a few could have been described as multi-site evaluation projects. In general, the reports consist of national studies, with different approaches, different theories of change, methods, instruments etc. The results of Workstream 2 give the impression that the European perpetrator programmes' evaluation landscape is scattered and complex, and meta-analytical conclusions about the work with perpetrators are hard to make. This situation results in the fact that the European specificity is not really taken under consideration in the international debates on evaluation research on perpetrator programmes; however, also “... existing evidence (e.g. from North American studies) cannot easily be generalised to a European context” (Hester et al., 2014, p. 3).

Therefore, the main goals of the project “IMPACT: Evaluation of European Perpetrator Programmes”, funded by the European Commission (Daphne III Programme), have been defined as follows:

- a) Filling the knowledge gap regarding evaluations of perpetrator programmes in Europe, and
- b) Offering solutions towards a harmonization of outcome monitoring (as a first step towards harmonization of evaluation methodology)

by means of:

- Providing an overview and analysis of the current practice of outcome monitoring in European perpetrator programmes (Workstream 1)
- Providing an overview and analysis of research studies evaluating perpetrator programmes (Workstream 2)
- Identifying possibilities and obstacles for multi-country European outcome research studies (Workstream 3)
- Developing a toolkit and good practice guidelines for outcome measurement (Workstream 4)

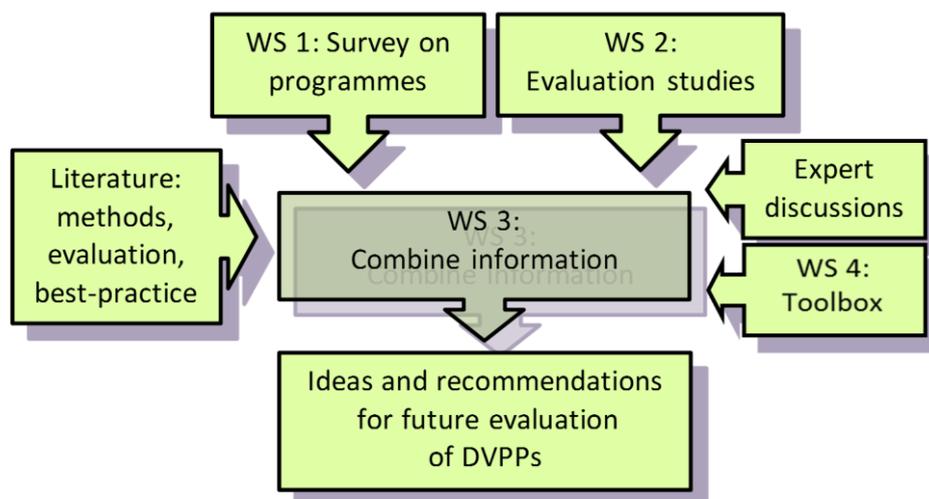
This working paper presents the methodology and results of the activities undertaken in Workstream 3. The identification of possibilities and obstacles for multi-site/multi-country European outcome research studies was based on:

- The (interim and final) results of Workstream 1: Overview of the practice of outcome monitoring (survey)

- The (interim and final) results of Workstream 2: Overview of the European evaluation studies
- Review of relevant literature in the field of research methodology and evaluation
- Interviews and discussions conducted with experts and practitioners from the area of work with perpetrators and cooperating services.

On the basis of the results of the survey in Workstream 1, the questions have been raised to what extent comparable methods and instruments for measuring outcomes or evaluation are already in use in different organisations, and which national and institutional prerequisites can be identified that promote or obstruct multi-national evaluation projects. Building on the knowledge generated by the review of research in Workstream 2, methodological elements for evaluation projects have been collected that show a high likelihood of transferability to various national conditions. In this sense, Workstream 3 deals with the feasibility of evaluation projects that include more than one site and more than one country. The activities in Workstream 3 have been connected to Workstream 4, in order to inform the development of the product “toolkit”, so that this product contributes to a harmonization of practices related to outcome measurement in programmes, and thus helps to prepare the ground for multi-country evaluation projects.

Figure 1. Position and function of Workstream 3 within the IMPACT-project.



Note. WS... Workstream. DVPPs... Domestic violence perpetrator programmes.

In this working paper, the work undertaken within Workstream 3 is summarized with a focus on the specific input by experts, and our conclusions for future evaluation projects in the field of work with perpetrators. More detailed analysis of the issues discussed in the working paper will be provided in two separate scientific articles.

The authors would like to thank all partners and associate partners of the IMPACT project, in particular Heinrich Geldschläger and Marianne Hester and their teams, for providing important interim results that helped move forward Workstream 3. We would also like to thank all experts who took part in the interviews and discussions.

2. Methods

The main aim of Workstream 3 was to identify possibilities and obstacles for multi-site/multi-country European evaluation projects on programmes in the field of work with male perpetrators of violence against women and children in close relationships (DVPPs – Domestic violence perpetrator programmes). The approach we used to identify evaluation models with a possible applicability in different European countries and contexts consisted of:

- (1) the analysis of existing literature on evaluation of perpetrator programmes (both in European and non-European contexts), as well as other relevant literature,
- (2) using the results of Workstream 1 and Workstream 2, together with additional analytical steps of the data that have been collected in these two Workstreams, and
- (3) conducting interviews and discussion with international experts and practitioners working with perpetrators of domestic violence.

It has to be emphasized that the focus of Workstream 3 was on (a) *outcome* (not on *process* or other aspects of assessing the work of perpetrator programmes) and (b) on *evaluation* and consequently on *research* projects, which goes beyond the measurement of outcomes in the daily practice of perpetrator programmes. (Nevertheless, these aspects are related and often have to be considered together: Outcome is dependent on the underlying processes; and ideally, the practices in measuring outcome are compatible or even part of the methodology which is used in evaluation projects.)

2.1 Analysis of literature

The findings from publications allowed for the preparation of an internal overview of the most common evaluation models, instruments, strategies and methods and supported the formulation of the questions for the expert focus group interviews, and expert discussions. Among others, important sources of information have been:

Batterer Intervention Systems: Issues, Outcomes, and Recommendations (Gondolf, 2002); *Domestic Violence Perpetrator Programs in Europe , Part I: A Survey of Current Practice; Part II: A Systematic Review of the State of Evidence* (Hamilton, Koehler & Lösel, 2013; Akoensi, Koehler, Lösel & Humphreys, 2013); *Changing Violent Men* (Dobash, Dobash, Cavanagh & Lewis, 2000); *Violent Men and Violent Context* (Dobash & Dobash, 1998); *Evaluating Criminal Justice Interventions for Domestic Violence* (Dobash & Dobash, 2000); *Rethinking Survey Research on Violence Against Women* (Johnson, 1998); *Description of a Voluntary Attendance Program for Abusive Men* (Payarola, 2011); *Interviewing Violent Men: Challenge or Compromise?* (Cavanagh & Lewis, 1996); *Men Researching Violent Men: Epistemologies, Ethics and Emotions in Qualitative Research* (Cowburn, 2013); *Encountering Violent Men: Strange and Familiar* (Gottzen, 2013); *Batterer Intervention Programs: A Report From the Field* (Price & Rosenbaum, 2009); *Counting Men: Qualitative Approaches to the Study of Men and Masculinities* (Patulny & Pini, 2013); *World Report on Violence and Health* (WHO, 2002); as well as two working papers from the IMPACT project: *Outcome Measurement in European Perpetrator Programmes: A Survey* (Geldschläger, Ginés, Nax & Ponce, 2014); *Overview and*

Analysis of Research Studies Evaluating European Perpetrator Programmes (Hester, Lilley, O'Prey & Budde, 2014).

Another important source of information for Workstream 3 was methodological literature on evaluation (e.g. Bamberger, Rugh & Mabry, 2006; Rossi, Lipsey & Freeman, 2004; Shaddish, Cook & Campbell, 2002; Stufflebeam & Shinkfield, 2007). Specific terms and glossaries from evaluation literature were provided for the project group in order to use a common terminology throughout the project.

In parallel to the IMPACT project, a significant evaluation project has been realized in the UK, called MIRABAL¹. This project has set out to answer some questions that also have relevance for the IMPACT project. The project partner Respect (UK) has a leading role in the MIRABAL project, therefore, the transfer of important developments in the two projects was given. Contacts and discussion among the two projects have been a valuable source of information (e.g. at the workshop in Bristol, UK, in 2013, where Nicole Westmarland presented recent developments and the methodology of the MIRABAL project). The publications to date that are related to the MIRABAL project have been a valuable source of 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). The results of the MIRABAL project will be presented in early 2015, and the project will certainly give new impulses for multi-site evaluation projects in the future. It is noteworthy that US-expert Edward Gondolf (2012) has taken the MIRABAL project as an example for a new trend in evaluation research, which he calls "practitioner-led evaluations" (p. 81) and which have set out to overcome some of the shortfalls of sometimes inappropriate academic approaches towards evaluating perpetrator programmes.

2.2. Applying the results of Workstream 1

The focus in Workstream 3 has been on evaluation models, i.e. designs, instruments and procedures of evaluations. Within Workstream 1, a survey was performed in order to collect information about European programmes in the area of work with perpetrators of domestic violence, as well as their practices concerning measuring outcomes and performing evaluation of their work. Consequently, we expected to gain information on promising practices that had already been in place or used in evaluation projects and that would show a certain degree of transferability. Although the authors state that "... the focus is not on scientific outcome research but on the day-to-day outcome monitoring perpetrator programmes regularly perform as part of their service delivery" (Geldschläger et al., 2014, p. 4), we expected to get hints from Workstream 1 for the planning of future evaluation studies. By going through the results of Workstream 1, we wanted to explore if there was a common denominator of elements that are already in place and approaches that could be chosen for future studies.

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

2.2.1. Important aspects from Workstream 1

In the survey of Workstream 1, data of 134 programmes from 22 countries were obtained, and the results of the survey have been published in a working paper (Geldschläger et al., 2014). The most important results and consequences for Workstream 3 are summarized below. For more detail, the reader is referred to Geldschläger et al. (2014).

Referrals

Most perpetrator programmes work with self-referred as well as agency-referred men. A portion of programmes excludes some kind of referrals or are concentrating on a certain kind of referral (22% of the programmes work with court-mandated men only; 19% do not work with court-referred men at all; Geldschläger et al., 2014, p. 12). Many programmes have a focus on court-referred men or self-referred (respectively referred by other agencies) men – but the majority of programmes work with various kinds of referrals. Future multi-site evaluation studies should keep this variety of referrals in mind.

Co-operation

Almost all programmes (98%) in the survey collaborate with some kind of network partner, among the most important: social services; police; victim protection; criminal courts; alcohol and substance abuse treatment services. 75% of the programmes say that they are part of an inter-institutional alliance against domestic violence, and a part of the programmes offer support for victims within their organisation. Thus, we can conclude that the programmes in the survey are following a “system approach” to a quite high degree, across a variety of different countries. The kind and quality of the networks might differ across countries and should be considered in multi-site evaluation designs. However, this finding also means that it is possible to concentrate on those programmes which are part of a coordinated network of institutions and services in the first place. Programmes that differ from that approach are of interest as well – they could be part of an evaluation programme and studied in a second phase, as they require different comparisons and designs.

Intake

Almost all programs (95%) have some kind of intake procedure: from one initial contact to longer individual counselling or assessment phases during the intake phase. This is basically a useful condition for realizing evaluation projects: For many programmes, the implementation of some kind of initial data collection is a “normal” procedure. Two thirds of the programmes use some kind of instruments at intake – but there are no prevailing instruments. A variety of questionnaires, manuals and other instruments are named. Here, some harmonization is needed – but it is also important that most of the projects are used to working with some instrument *in principal*. Evaluation studies in the future can build upon this competency.

The majority of the programmes can apply some kind of inclusion/exclusion criteria, based on the intake procedure (e.g. three thirds name “alcohol/substance abuse” and “mental disorders” as excluding conditions). This means that the majority of the programmes can make statements about characteristics of their clients regarding certain behavioural and

psychological characteristics. Again, this is an important finding in terms of multi-site evaluation studies: It seems feasible to implement data collection procedures before the programmes start, in order to study differences between certain kinds of participants and drop-offs, respectively, later on.

Setting

Around two thirds of the programmes apply group work (eventually mixed with other kinds of intervention); individual counselling is also frequent but group work is the predominant way of intervention. For multi-site evaluation studies, both approaches (group work, individual setting) seem interesting. Other approaches are less frequent (e.g. couple counselling, mediation); these approaches could be included into certain evaluation designs, but they would require specific evaluation models.

Concerning the “dose” of treatment, there is a lot of variation across programmes, with the majority of programmes in the area of 14 to 52 sessions, and 14 to 52 weeks duration. This variability is important for evaluation studies in terms of a dose-response approach, meaning that longer and shorter interventions could be compared.

Working approach

Concerning the working approach of the programmes for their concrete work with perpetrators, most programmes are in the area of Cognitive-Behavioural Therapy (CBT) and a psycho-educational approach. Together with combinations of CBT, psycho-educational or a Duluth-like approach, two thirds to three fourths of the programmes have such a basis. These approaches are, of course, not identical – but they have some elements in common, especially on a theoretical level (learning theory). For the US, Gondolf (2012) speaks of two major categories of approaches in the field of work with perpetrators: c-behavioural and psychodynamic approaches (respectively the “new psychology”-approach), each approach being very diversified in itself. In the European case, we see a variety of approaches, sometimes combining elements from different approaches.

Measuring outcomes

As far as outcome measurements are concerned, 81% of the programmes say that they measure the outcomes in some way (p. 33). The most important sources of information are the clients themselves (94% of those who said that they measured their outcomes), facilitators (63%), and ex-/partners (60%), but also other services (54%), police/court reports (43%) and victim support services (32%) were included. Follow-ups were reported by two thirds of the programmes.

This result shows that integrating other sources of information than the men on the programmes is a frequent and common practice throughout the projects, which future evaluation studies can build on. The *perspective of the partners* is seen as the central source of information in *evaluation* projects, and in most of the countries, programmes can be found which include information by the ex-/partners. According to the information of the programmes in the survey there are some countries, like the UK, where this seems to be common, and others where it may not be common (e.g. France, Switzerland). Also, a connection to *programmes within the Criminal Justice System* in some countries must be

kept in mind, namely that “... ‘mainly court-mandated’ programmes tend not to use information from partners and victim support services, as expected, because a lot of these programmes indicated that they don’t contact partners ... and don’t measure outcome with partners ...” (Geldschläger et al. 2014, p. 46).

However, we find programmes in 17 countries, all over the continent, that include the *ex-/partners’* information in measuring outcomes. We therefore conclude that there should be no principal problem with a multi-country-study that includes the partners in assessing the outcomes of programmes. Where this is not possible, the respective programmes will play a different role in a future evaluation strategy.

Regarding the *men on the programmes*, the most frequent outcomes that were measured were the following: Non-Violence / decrease of violence, attitudes and beliefs, client’s communication skills, and decrease in risk of violence. Psychological aspects, fathering and quality of life were measured as well, by between half to two thirds of the programmes.

The most important dimensions that the *ex-/partners* were asked about were: Violence, feelings of safety, decrease in risk of violence, children’s safety, quality of life, man’s attitudes, man’s communication skills, man’s fathering skills.

All in all, there is a high variability of methods how to measure outcomes of perpetrator programmes. “One of the main conclusions of this survey is therefore that European perpetrator programmes are far from a consensus about standard methods of outcome measurement and a need for some harmonization seems quite apparent.” (Geldschläger et al., 2014, p. 56) In Workstream 4 of the IMPACT project, we take a step into this direction.

2.2.2. Conclusion

We conclude from the results of the survey in Workstream 1 that there are no principal problems to measure outcomes in many countries and for many programmes in a comprehensive way, including the perspectives of men on programmes, partners and children, official records, facilitators – however, additional resources are needed to join an *evaluation* project, for most of the programmes.² Practitioners seem open to evaluation, and various data-collecting procedures are in use. Nevertheless, evaluation studies go far beyond the measurement and monitoring of outcomes. Consequently, multi-site evaluation studies should be performed as separate projects with additional funding. Based on the findings of Workstream 1, we do not consider it feasible to perform multi-site/multi-country evaluation studies that are based on the ongoing practices of the programmes, regarding outcome measurement or collection of information in the intake phase. The existing practices are simply too heterogeneous to make use of it in a multi-site context. Additional procedures in evaluation studies will require additional resources for practically all programmes to be involved.

² This is also true for measuring outcomes of perpetrator programmes: as far as obstacles for outcome measurement are concerned, the main problems for a lot of programmes are constraints in time and resources. A smaller part of the programmes (approx. one third) see “lack of methodology” and “legal/institutional context” as obstacles for measuring outcomes (for legal obstacles see above: programmes in the Criminal Justice System in some countries). However, almost all programmes in the survey are interested in improving their outcome measurements, in evaluation and international exchange. Therefore, many practitioners want more human resources, economic resources, methods, resources for data analysis, guidelines/recommendations, and training/consultation.

2.3. Applying the results of Workstream 2

In Workstream 2 the focus was shifted from practice to research. The aim was to gain knowledge about the approaches used in evaluation research studies on perpetrator programmes in Europe. The research team identified studies published in scientific journals as well as “grey literature”, both in English and other European languages. Methods, designs, measures of outcome and further aspects of the studies were analyzed, so the results of Workstream 2 were of high importance for the goal of Workstream 3 (i.e. to give recommendations for future evaluation studies regarding evaluation models).

2.3.1. Important aspects from Workstream 2

Frequencies by country

Hester et al. (2014) describe in detail their approach towards identifying existing studies. Four meta-analyses and syntheses were found which were based on fifteen studies in total. Interestingly, most of these studies come from the UK (7) and Spain (5), and these two countries were also among the ones with the most programmes identified in Workstream 1. Together with further steps to identify evaluation studies both published and unpublished (or published as grey literature), a total of 65 studies could be used for further analysis in Workstream 2. Again, many of these studies come from Spain and the UK, as well as from central European countries³. We can see an overlap of frequencies regarding *programmes* (Workstream 1) and *evaluation studies* (Workstream 2) which makes sense: Countries with more (visible) perpetrator programmes tend to perform and publish more evaluation studies.

Problems related to the studies

There was a range of problems in the analyzed evaluation studies. Some of the most important are listed below:

- *Design:* As regards research design, there were almost no Randomized Controlled Trials (RCTs) to be found. Of n=58 studies from the five European regions that addressed *outcome* (they were focussed on outcome alone, or on outcome and process), only 2 had a RCT-design, and these were studies conducted in a prison and in a psychiatric clinic.
- *Control groups:* Many studies lack a control group, or the control groups are small. This fact is a major limitation of the available studies as control groups (or equivalent elements of control that allow causal inference) are crucial for the internal validity of any study.
- *Terminology:* In the studies, a range of definitions of crucial terms could be observed (e.g. “start”, “completion”). A consistent terminology does not exist.

³ For their analyses, Hester et al. (2014) used a classification of countries into five regions where evaluation studies were identified: West (UK, Ireland), South (Spain, Portugal), North (Sweden, Denmark, Finland, Iceland), East (Croatia), and Central (Germany, Austria, Switzerland, Netherlands).

- *Points of time:* There were differences in approaches and methodologies as regards the points of time when data were collected. Hester et al. (2014) have introduced a scheme for their analysis that should be used and promoted for a common understanding, concerning points of time of observation in evaluation studies. It consists of five points of time at which information should be gathered in an evaluation (T0: Pre-Start; T1: Start; T2: During intervention; T3: End; T4: Follow-up). This scheme is described further below.
- *Attrition:* Attrition rates were unclear, or it was unclear in which phases attrition occurred. Information on sample sizes and attrition at every stage of the intervention process was missing in most cases. The information given varies a lot, concerning samples and sample sizes – due to a lack of standards and established terminology for evaluation studies on work with perpetrators programmes.
- *Referral:* It is essential to have information about the way of referral or programme entry pathways (self-referred or court mandated attendance, other ways of referrals). Also the reasons for dropping out, socio-demographic and other information of all subsamples (clients being referred before the training starts; at the beginning of the training; drop-offs; completers) are needed. In many cases, this information is missing.

Measuring outcomes and impact

Regarding outcomes, self-reported data from perpetrators, official data from the criminal justice system and data from the victims-survivors were used in all regions, with the exception of the South, where data from the partners did not appear in the available studies. However, there are programmes in the Southern region which include the victims' perspectives, as we know from the analysis in Workstream 1. However, there seems to be a difference of the practices and legal possibilities of programmes within jails and community-based programmes. Programmes within the criminal justice system do not integrate the partners' perspectives, however they are evaluated more often than community-based programmes. Also, they use a lot of validated/standardized psychometric instruments, focussing on psychological aspects and changes within male perpetrators. This reflects a strong psychological approach towards the evaluation of perpetrator programmes in the criminal justice system in the Southern region, with the most studies coming from Spain.

The most comprehensive outcome measures have been found in the Western region, "... including change in attitudes towards women and violence against women, official and self-reported recidivism, repeat victimisation, the quality of life, safety and well-being of the women/victims and their children, levels of parenting stress and drug / alcohol use of the perpetrator." (Hester et al., 2014, p.26) Risk assessment instruments are found in studies in the Western and Southern regions.

In general, the dimensions that are studied and the instruments that are used vary a lot from region to region (and from study to study). This lack of a consistent methodology makes it difficult or even impossible to compare the studies.

Thus, when it comes to giving an overall statement of the effectiveness of perpetrator programmes on the basis of the available evaluation studies, only a very general picture can be provided: there are either some statistically significant findings that point into the

direction of positive changes; or there are promising results that are not statistically significant. This pattern was found for a variety of dimensions (behavioural change; attitudinal change; psychological change; change in risk factors; motivation and adherence). However, there are limitations of the findings concerning impact, namely small sample sizes in many cases, high levels of attrition (or information on attrition missing), and an over-reliance on men's self-reported data in the Southern region.

2.3.2. Conclusion

Regarding Workstream 3, despite of all the problems found in the studies, the results point into the direction that it is possible to get and combine various data in all regions. There are different approaches regarding evaluation but it seems feasible to apply certain instruments and to integrate certain sources of information across countries (with some exceptions, e.g. programmes in jails, which have to play a different role in a future evaluation research strategy).

The main problem today seems to be the inconsistent methodology regarding the evaluation of perpetrator programmes, which makes it difficult to compare the studies in different respects (designs and control groups, terminology, instruments, sources of data). A harmonization of the evaluation methodology is needed in order to overcome the fragmentarization of evaluation research on perpetrator programmes. This seems possible, as the authors of the working paper on Workstream 2 conclude:

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

2.4. Experts' focus groups interviews

As far as evaluation studies of perpetrator programmes are concerned, there is a variety of stakeholders to be identified, all of them with their interests and perspectives, which can differ from each other but may also overlap:

- *Victims-survivors of domestic violence:* The female partners and children of violent men will want to know if a programme is effective to stop their partners'/fathers' violence and to increase their safety.
- *Practitioners in victim support:* Professionals in victim support organisations want to know if the programmes are effective and whether they are effective enough to justify the use of additional financial resources that otherwise could be used for increasing victim support.
- *Practitioners in the intervention chain:* Police, prosecutors, courts, probation officers and other stakeholders are also interested to see if DVPPs have a specific effect (in addition to the existing interventions).
- *Practitioners in DVPPs:* These professionals will be interested in their works' results and will want to know specifically whether or not their work is effective and violence

decreases, and if so, how they can improve their work into the direction of even better outcomes.

- *Programme managers of DVPPs*: They will be interested in the effects of the programmes as well as in improving the quality of the programmes, i.e. their interest is to demonstrate effectiveness and improve outcomes.
- *The public*: In general, the public may be interested in reducing domestic violence in the society and to increase safety. The public discussion on the topic often takes place in the media who have their own logic and interests. Quality media may raise questions similar to the other groups of stakeholders, while other segments of the media work with shocking, flashy and extraordinary contents; often, a call for harder punishment of perpetrators and stricter laws are heard from this side of the media, however there is some interest in the effects of alternatives like DVPPs.
- *Funders, politicians*: Stakeholders of this category may want to demonstrate to the public their activities against violence, at the same time to make sure that the funds they are assigning to programmes have effects and lead to the fulfilment of pre-defined goals; at the same time, they may wish to optimize the usage of resources, i.e. they are interested in answering questions regarding cost-effectiveness.

In the *expert focus group interviews* and in the *expert discussion* (see below), the issue of different interests as well as overlapping interests of the various groups of stakeholders was an issue.

The **expert focus group interviews** (a social research method that combines expert interview and focus group interview; see Babbie, 2003; Silverman, 2013) were conducted in October 2013 during the project's *Workshop A*, organized at the University of Bristol, UK. Four 45-minutes focus group interviews were carried out with 18 experts (researchers and practitioners working with perpetrators) from 11 European countries (Austria, Czech Republic, Denmark, Finland, Germany, Ireland, Poland, Portugal, Norway, Spain, United Kingdom). In each group, the facilitator (a member of the project team) asked three main questions:

- Which fostering and hindering conditions regarding feasibility of a multi-site-design in different countries do you see, on a national or local level?
- Which fostering and hindering conditions regarding feasibility of a multi-site-design in different countries do you see, on an international level?
- Which possibilities and obstacles do you see, concerning the following idea: „To provide a tool for practitioners (in perpetrator programmes), to collect comparable data across countries; pooling these data, analyse centrally“?

After the interviews, the facilitators prepared internal reports, which served as a base for the presentation during Workshop B as well as for further analysis (see below).

The **expert discussion** took place during Workshop B organized in April 2014 in Copenhagen, Denmark. 41 practitioners, researchers and policy makers from Austria, Bulgaria, Czech Republic, Croatia, Denmark, Estonia, Finland, Germany, Ireland, Italy, Lithuania, the Netherlands, Russia, Spain, Sweden, Switzerland and the UK discussed the issues and questions provided in the presentation *Possibilities of Evaluation in a Multi-Site/-Country Approach*:

- What possibilities and obstacles do I see in my organisation?
- Which questions should be addressed in evaluation studies:
 - Does it work?
 - To what extent / how effective?
 - What works?
 - Specific questions (Who re-assaults, who needs what etc.)

The results of the discussions conducted in four groups were collected and served as a base for further analysis (see below). Further feedback on the concept of multi-site/multi-country evaluation studies on DVPPs was collected at the project's final conference in Barcelona, Spain (October, 2014).

3. Results: Experts' perspectives

The results of the expert focus group interviews as well as the expert discussions were summarized as a collection of issues, which have been a source of information on how to move on with the development of a methodology for evaluating perpetrator programmes. These issues have been categorized into five main groups: questions, possibilities, advantages, challenges and obstacles.

3.1. Questions

One of the most important and common questions which appeared during the interviews and discussions referred to the fundamental understanding of "perpetrator programmes". According to the experts in various countries, the definition of such a programme may vary: Apart from typical programmes (i.e. regular group meetings, conducted by professional facilitators, structured, co-ordinated, evaluated) the following interventions were named as elements in or around the work with perpetrators of violence in close relationships: phone help-lines; individual settings; informal, voluntary community or/and church meetings; meetings of Anonymous Alcoholics etc. Therefore, one of the first steps in thinking about multi-country European evaluation research on work with perpetrators would be to give a common but also relatively broad **definition of a perpetrator programme**.

Another important question addressed the issue of **similarity of programmes**. Most experts shared the opinion that only outcomes from similar programmes (same approach, similar working methods, similar character, length, size etc.) should be compared. This assumption, however, caused doubts regarding the (insufficient) number of European programmes which were similar enough to be compared.

The issue of similarity was followed by the question on the **type of programme** to be studied. Some experts underlined that the future study should focus on one type only: either *mandatory* programmes or *voluntary*, while others proposed to study both approaches or to concentrate on "mixed" programmes (with both kinds of referral pathways in the same programme).

Finally, the experts addressed the issue of the **research procedure** of a future study. Concrete questions prevailed, such as who would be the researchers, who should collect the data etc. The pros and cons of various possibilities were discussed, e.g. to have either independent, external researchers, or to have practitioners and persons directly working with perpetrators to collect data and perhaps to do analytical steps. Motivation and genuine interest in evaluating the work with perpetrators may be an advantage in the latter case, but conflict of interest may be a disadvantage. According to the experts, both ways have their pros and cons (different access to the data and sample; costs; different level of analytical skills; the issue of ‘objectivity’; time etc.), and the decision which procedure to choose should be considered well.

3.2. Possibilities

Many experts expressed their interest in performing and supporting some kind of multi-country European outcome study. Consequently, the question about possibilities for such studies was discussed the most. In the related discussions, **standards for programmes** as well as **outcome measurements** to be used in evaluation studies were of particular interest to many experts. A starting point for defining minimum standards for perpetrator programmes as well as for their outcome measures were seen in the *Convention on Preventing and Combating Violence Against Women and Domestic Violence* (“Istanbul Convention”, Council of Europe, 2011), in the briefing note on outcome criteria by Westmarland, Kelly and Chalder-Mills (2010), or in the *Guidelines to Develop Standards for Programmes Working With Male Perpetrators of Domestic Violence* (Work with Perpetrators [WWP] – European Network, 2008). It should also be underlined that in the experts’ opinion the standards should be based on gender equality politics.

The ‘minimum’ term was also mentioned in connection to the research and evaluation studies, in the sense that a **minimum of standards** is needed **in each country-specific study** or each part of the study, to be able to “pool” the different parts. Several elements were named that have already been in use at various occasions, e.g. the ecological model; stage observation model (pre, post, follow-up); certain instruments (e.g. Violence and Abuse Inventory). Researchers have already started to describe and analyse various relevant issues, e.g. European and national context (legislation, referral way etc.), types of perpetrators, frequencies and types of violence, desirable areas of change (perpetrators’ violent behaviour; communication; opinions and identities; family relationships etc.). The discussion led some experts to raise first research questions (without providing answers at that point of the discussion) on how to address these questions):

- What makes the greatest contribution to end violence and increase victims’ safety?
- Do men change? Why? How?
- What works for whom?
- What (other) factors affect perpetrators’ lives (class, religion, non-/migration background etc.)?

Finally, the experts addressed the conditions which may help to increase the **quality of future multi-site/multi-country evaluation studies**, e.g.: triangulation of research methods (qualitative, quantitative, observations); triangulation of data and theories; taking under consideration perpetrators', victims' and practitioners' perspectives. Moreover, in the experts' opinion, the involvement of umbrella organisations and local/specialized NGOs (migrants groups, LGBTQ groups, religious groups) in the research process can also be highly useful.

3.3. Advantages

In all interviews and discussions, the experts underlined the advantages of multi-site evaluation studies. A **good** and **useful** way of measuring violence and other **outcome measures** would have to be used or developed in such a study, which would not only be beneficial for practitioners from all European countries but would also contribute to filling in the gap in knowledge and research on violence in general: according to some experts, the existing methodology was insufficient. Especially practitioners expressed the hope that evaluation might support the **continuation of the existing programmes**, as some decision making bodies connect further funding to positive results of programme evaluation.

Moreover, as the experts pointed out, the idea of integrating the perspective of victims and victim organisations would lead to **more comprehensive evaluation studies**, including the **broader system** that was involved in the intervention. Multi-country European evaluation studies could also contribute to the **promotion** of unique national perpetrator programmes that are not well-known yet among the various stakeholders, and the results regarding the programmes would serve as a **knowledge base** for practitioners who are willing to improve the quality of their work, avoid common mistakes etc. Moreover, as an additional effect, the attention given to an international study on work with perpetrators will increase the **social awareness** about the existence and role of perpetrator programmes in certain countries, and, according to some experts, this will help with more effective promotion both on national and international level. If a multi-country European evaluation study was performed as a European project, it might be the only chance for evaluation of programmes in those countries where programmes have not been evaluated so far, due to a lack of resources and methodology.

3.4. Challenges

The experts agreed upon the **recruitment process** as one of the biggest challenges regarding multi-country evaluation studies, especially in case of no additional funding. They underlined that such a study could only be successful if the researchers could convince practitioners, programmes' staff, and other key actors to sacrifice **time** and **resources** to take part in the research. In their opinion, arguments about filling the knowledge gap in the area of perpetrator programme evaluation studies might not be enough to motivate key actors to participate. Therefore, other **benefits** should also be emphasized that result from participating in evaluation research studies, such as external training for the programmes' staff (not only on evaluation methods), study visits, certification (e.g. done by Work With

Perpetrators - European Network, or national agencies such as Respect, UK, or BAG Täterarbeit HG, Germany). According to the researchers among the experts, one of the most important benefits of an evaluation study will be the feedback from the clients, which may lead to an increase of the programme's quality and effectiveness. In addition, programmes participating in the study might have a better chance to receive funding in the future when an evaluation process has been completed and the results prove that the programme has some impact. The European dimension of an externally led evaluation study would increase credibility of the results on the national level.

Another challenge is to find an **agreement** regarding the usage of the same **evaluation models** (criteria, designs, methods, instruments). As the programmes and practical methods are different in many countries, different criteria and instruments are used to measure the outcomes at this point in time. Therefore, the challenge is not only to select and use the same tools but also to select the most universal and transferrable ones.

3.5. Obstacles

Regarding to the experts, carrying out multi-country European outcome research may encounter a number of obstacles. One of the most serious problems is **funding** – as conducting such a complex study will generate a lot of costs. Therefore, finding the financial support for such a project is crucial but also highly challenging. Funding is needed not only for covering the work of a team of external researchers and other project expenses (travel costs, equipment etc.) but also for covering the extra work of the programmes' staff, especially in the situation where practitioners are underpaid or even work on voluntary basis (which is often the case with programmes located in post-socialist countries), and/or where evaluation of the programme is not included in its regional or national funding.

The **different context** in different European countries is another obstacle which influences both the programmes' character and the future research process. Different legal contexts may influence the choice of specific methods and basic assumptions of a research project. Legal regulations regarding the intervention system, personal data protection etc. may prevent researchers from **retrieving certain data** in certain contexts. One of the common examples of such a situation was given by an expert from Spain, where the staff from the perpetrator programmes in the justice system does not have access to the victims, and consequently, inclusion of the partners' perspective might be difficult or even impossible to achieve.

Another issue is the **cultural context** of different European countries which may also influence the research process; in some countries, the **understanding of crucial terms** such as *domestic violence, victims' rights, inappropriate behaviour, gender roles, gender equality* etc. may differ and therefore, designing multi-country research tools can be highly challenging. Moreover, countries vary in terms of their **economic** situation, the role of **religion(s)** (the degree to which politics are influenced by religious ideas, more or less secular countries), the variety of cultural backgrounds in connection to **migration** etc., but also in terms of the degree of **implementation** of **gender politics** and gender mainstreaming,

the degree of implementation and development of *perpetrator programmes*, as well as the level of *acceptance of the importance* of such an intervention system.

In terms of the programmes, the experts see important differences across countries in terms of *theories of change, criteria of success, types of men* who might take part in the perpetrator programmes, and different approaches regarding the *clients' gender* in the groups: There are programmes that work with men only, others have a coeducational approach, with different groups/paths for male and female perpetrators or with no differentiation. To summarize, in the experts' opinion the most serious (methodological) obstacle for multi-country European evaluation research is the fact that *perpetrator programmes*, for various reasons, are simply very *different*.

Finally, some experts doubted that it is possible to conduct a complex European study as there are still not enough perpetrator programmes (in some countries, like Hungary, programmes do not exist yet; Kutrovátz, 2013), which could result in a serious sampling problem.

3.6. What results from the expert focus groups and the discussion

The contributions of the experts to the group discussions and their answers to various questions were summarized to the following expert recommendations:

- In the first place, a *definition* of “perpetrator programmes” must be given that should be considered in multi-country evaluation studies. The standard programmes (co-operation with victim services and other agencies; based on CBT or a psycho-educative approach, with a gender focus) seem the most appropriate at this point, also for methodological reasons: we need similar programmes in different contexts, and not different programmes in different contexts. The programmes must also be able to collect the appropriate outcome data, which is dependent on cooperation, i.e. a coordinated approach. Moreover, the focus of the study should be on programmes which are *compatible* to existing *standards* and *policies* (e.g. the Istanbul Convention).
- *Research project*: Working together in a new, funded project was seen as the most feasible way to realize a multi-site/multi-country evaluation project. The initial idea of using the toolkit for evaluation by collecting and analysing data centrally was questioned - at least at this point of the process. Nevertheless, the toolkit should be implemented broadly, research should build upon the toolkit as far as possible, and at a later stage, it might be possible to go back to the initial idea. The difference of the toolkit (as a resulting product of Workstream 4, to measure outcomes) and instruments as part of an evaluation methodology was emphasized by the experts. Including evaluation projects into the daily work seemed challenging.
- Future studies should be designed in a rather *comprehensive way*, including the *system* that is involved in the intervention. Social impact of perpetrator programmes could extend the basic design about the impact of the programmes. Different *context* in the different countries should be considered from the beginning, in terms of legal regulations, aspects of the intervention system, and cultural issues.

- Evaluation criteria, designs, methods, instruments etc. should relate to existing **good practice in evaluation**, and take the views of the **experts** into account.
- The many **questions** to be answered by evaluation studies should be **systematized**, preferably by relating them to certain groups of **stakeholders**.
- Evaluation models should show a high degree of **transferability**, to be able to include many projects in many countries, in the long run.
- Concerning the type of programme: **referral pathways** (self-referred, agency-referred, mixed) could be used as a variable, not as a selection criterion.
- In any case, **benefits** resulting from the participation in evaluation studies should be emphasized in the recruiting process.

4. How to move on with the evaluation of DVPPs

4.1. Starting point: Where we are at the moment

As has become clear from studying available sources on evaluation of DVPPs as well as the experts' perspectives, the initial idea of Workstream 3 has to be modified. It has not turned out as feasible to plan, design and perform big evaluation studies that answer the various questions put forward by the different stakeholders in a comprehensive way, at least not at the moment. Rather, we have developed a concept for co-ordinated evaluation efforts, to be seen as a series of smaller evaluation projects with a similar methodology and a possibility for the co-ordination of these projects.

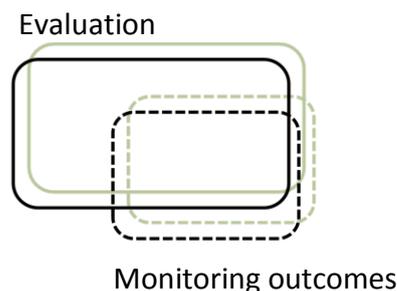
To summarize, the prerequisites for multi-country evaluation/multi-site projects are the following:

The **main motivation** for such a project is to make use of the high variability of national conditions and "systems" (practices of institutions, victim support, DVPP etc.) throughout Europe. This variability can be a resource for evaluation studies, to study DVPPs in different contexts. Furthermore, as many programmes have not been evaluated yet, multi-country/multi-site evaluation studies can be an opportunity to broaden the empirical basis, concerning DVPP. As outlined above, there is a high diversity of work with perpetrators over countries, but also similar programmes can be found (e.g. CBT- or Duluth-based DVPPs), and elements of co-ordinated community responses (CCRs) are widespread. Procedures to measure outcomes are in place in many countries, but there is a high methodological variation, which makes the existing studies difficult to compare. There are differences and problems regarding terminology and concepts (e.g. "completion", "success", etc.) as well as designs (e.g. concerning the points of time of measurement and the source of information); control groups are difficult to establish, and there are ethical concerns around control groups; attrition is a major problem in practice, concerning men on programmes as well as victim-survivors. However, as Hester et al. (2014) conclude from their compilation of studies, "... it would be possible in the main to take elements from different approaches in order to start developing a robust evaluation methodology." (Hester et al., 2014, p. 39)

Experts have found it worthwhile to follow the idea of a multi-country approach, and the different groups of stakeholders are interested in gaining answers to the question of the

programmes' effectiveness of various kinds of DVPPs. As various stakeholders have argued, there should be a focus on programmes which are compatible to existing standards and policies (e.g. Istanbul Convention). However, it became clear that evaluation studies that are able to answer the relevant questions require a lot of effort and resources, and project managers and professionals have been sceptical about integrating that kind of work into their daily routines. In other words, it did not seem feasible to head into the direction of international, multi-country/multi-site-evaluation studies without additional resources. Stakeholders in practice may be convinced to change from instruments which they are using for assessing their outcomes to alternative instruments – but a number of problems remain in terms of evaluating the work with perpetrators, and these problems cannot be solved within the day-to-day-practices of the programmes. It is important to acknowledge that sound evaluation studies are broader in their concept and that they need more resources than measuring outcomes. The necessary work and efforts in evaluation projects go far beyond the daily routines within outcome measurement, although measuring outcomes can be a good start and basis that makes DVPPs “research-ready” (see Gondolf, 2012; see figure 2). Consequently, practitioners emphasized the need of additional funding if they were to participate in evaluation studies.

Figure 2. Overlapping and specific aspects of (a) evaluation studies and (b) practices to monitor outcomes.



All in all, we can conclude that a basis for multi-country evaluation projects is given – but a harmonization of the methodology is definitely needed. The DVPPs and the victim support services are needed for evaluation projects; thus, to enhance the chances for success, future evaluation studies should be planned as separate projects with sufficient funding to cover the additional costs of the organisations that will be involved.

4.2. Where we should head towards

Ideally, in some years ahead, the fragmented situation in evaluation of DVPPs is overcome, by a harmonization of the evaluation methodology, so that the results of various studies can be compared and complement each other. The idea is to have a co-ordinating institution which is promoting, supporting and monitoring future evaluation projects of DVPPs, in terms of designs, methods, instruments, as well as overseeing which specific questions should be addressed and answered. This will be not a full *DVPP-evaluation research programme*, as it is not realistic that a series of studies over several years can be anticipated and planned in a comprehensive manner, mainly due to restrictions in funding. However, it is very likely that

there will be various evaluation studies, some on a national level, others on an international level (e.g. as EU-projects) – and these studies need to be comparable, and it should be possible to link them in the manner of a “jigsaw-puzzle”.

We propose to use the *Work-With-Perpetrators-European Network* (WWP-EN) as a co-ordinating institution, supported by the *WAVE-Network*, for several reasons: WWP-EN has taken the leading role in linking DVPPs in the EU and beyond. As an umbrella organisation of DVPPs together with co-operating victim support services, steps into the direction of standards for the concrete work with perpetrators have already been taken, and are likely to be taken in the future. Evaluating this work is just another logical step – as is taking care of **standards** how to do this **evaluation** work (see below).

This corresponds to what Gondolf has called “practitioner-led evaluations” (2012, p. 81), respectively evaluation projects that are initiated or co-designed or influenced by the practical field. In many cases, as Gondolf has pointed out, the co-operation of researchers and practitioners has proven as difficult, each side only partly understanding the other, with much potential for improvement.

4.3. Standards for evaluation projects

As we have learned from Workstreams 1 and 2 as well as from specific literature, methodological standards and a common terminology seem crucial for a meaningful development of evaluation projects in the future. Evaluating work with perpetrators in general and DVPPs in particular is a very specific task, and procedures from other areas (e.g. evaluation of psychotherapy) cannot be applied in a “copy & paste”-manner. Some of the specific aspects of evaluating DVPPs are:

- *Concepts and terminology:* There have been extensive discussions up to now what DVPPs should head for, with consequences for terms and concepts such as “success”. There are various groups of stakeholders that have expressed their positions and interests concerning the outcomes of DVPPs. These positions must be taken into consideration when evaluation studies are planned.
- *Source of information:* Men’s information is considered as not sufficient by experts in the field (see e.g. Gondolf, 2012). Partners’ information, police or court records are needed to complement the information needed. Partners’ information is seen as most reliable and most meaningful, and thus should be the central indicators. However, this in turn raises ethical questions that have to be considered (see Downes, Kelly & Westmarland, 2014).
- *Observation times:* In order to deal with the relatively high levels of attrition (regarding perpetrators as well as victim-survivors) on the one hand, and to assess the sustainability of the changes achieved by the programmes on the other hand, a simple pre-post-design is not sufficient for evaluating DVPPs. Hester et al. (2014) have proposed a scheme of points of time for observation to be used in this context. They have defined four points of time: T0, when a man is referred to the programme; T1, when the programme starts; T2: during the programme; T3: at the end of the programme; T4: Follow-up, e.g. 6 months after the programme has finished. This

scheme, together with requirements concerning the information to be collected at each point of time (who has to be asked, what instruments should be used), can constitute a central standard for the evaluation of DVPPs. Together with a sufficient description of who the men are at T0 (e.g. concerning socio-demographic or psychological characteristics), and what the way of referral to the programme has been, the problem of attrition can be studied in a much better way than today.

4.4. Promotion of standards

Once such methodological standards have been developed, they have to be promoted and used in evaluation studies. Within Workstream 3, various ways of promoting such standards were discussed. The following possibilities have been identified:

- A task force within the WWP-EN, responsible for evaluation issues, could play the key role in harmonizing evaluation methodology, collecting best-practice in DVPP-evaluation research and “steering” series of evaluation studies. Future evaluation projects will need to co-operate with the programmes in the field, as it has been done in the past. Therefore, it is likely that evaluation projects are based on member organisations of WWP-EN, which will be sensitive in terms of the evaluation standards of WWP-EN.
- In the same way, international networks of victim support organisations, especially the *WAVE-Network*, will play a key role. Evaluation of DVPPs need the perspective of the victims – thus, such projects will be in contact and co-operating not only with WWP-EN member organisations but also with WAVE member organisations. As WAVE and WWP-EN are co-operating, these two European umbrella organisations can have a significant influence on the harmonization of evaluation methodology, via their member organisations. Moreover, there is the chance to initiate (national and international) evaluation studies within WWP-EN. Especially concerning multi-country EU-projects, WWP-EN is the ideal European organisation to co-ordinate proposals in the field of evaluation research.
- Promotion of the methodological standards by an EU-funded evaluation project, as a next milestone. The concrete methodology can be developed within this future project on the basis of the feasibility assessment of Workstream 3, with a pool of seemingly suitable programmes in various countries, to be identified by Workstream 1 data and the WWP-EN. Once a project has set out to make a first multi-country step in the evaluation of DVPPs, others can follow and will build upon the experiences of the pilot project.
- Promotion of research-like practices by the toolkit of Workstream 4. Although the toolkit is about outcome monitoring in the daily practice of the WWP-programmes, it is a step towards harmonization of collecting data, where future studies can build upon in various countries. When more programmes than today will use the toolkit, they will integrate research-like routines into their daily work that will prepare the ground for an evaluation project. Outcome monitoring will make programme teams “prepared for research”.
- EU-programmes that harmonise research and practice across the EU are also possibilities of harmonizing methods in evaluation research. There are EU-

programmes that work towards developing a European Research Area, e.g. COST. On the level of practice, the networks WWP-EN and WAVE will play a leading role in promoting methodological standards in evaluation, when they help to choose programmes and encourage their members to take part in multi-country evaluation studies.

4.5. A comprehensive approach towards evaluation of DVPPs and the role of multi-country-studies

To start from the “jigsaw-puzzle”-metaphor from above, different kinds of evaluation studies can answer different questions – the important point is to make sure that these studies can be compared and linked to each other. This can be ensured by the use of the same methodology (or at least a core of comparable procedures and instruments). We propose to look at the various kinds of evaluation studies from the perspective of the categorization outlined below. (More in-depth information on this and other aspects will be provided in separate articles to be written and published as part of the IMPACT project.)

4.5.1. Basic single-context studies

These are studies that evaluate the impact of a DVPP as part of a CCR. Such studies are performed in the field, not under laboratory conditions, and consequently, they are more likely to apply quasi-experimental designs because random assignment of research units (e.g. men) to treatment conditions is often not possible, for methodological and ethical reasons. Drop-offs have been used as control groups, or other groups of men who did not complete the interventions (e.g. men who were excluded from the programmes or not admitted to the programmes) are compared to those who have completed programmes.

We want to propose three elements here, as part of a future evaluation methodology, to be refined e.g. by the *WWP-EN evaluation task force*: (a) observations / points of time; (b) indicators of success; (c) design.

(a) Observations / points of time

Hester et al. (2014) have provided a scheme in their analysis of European evaluation studies to describe “shape and size of the sample at the different stages of the evaluation process” (p. 17). This scheme should be used in evaluation studies on DVPPs, to define the different points of time of observation. The scheme consists of five points of time, together with proposals for sources of information at the respective point of time:

- T0: „Pre-start“. It is important to know what kinds of men are referred to the programmes (socio-demographic data, psychological characteristics, substance misuse, way of referral etc.), and whether or not they enter the programmes. The comparison of the samples at T0 and T1, together with the reasons for dropping off, can provide meaningful information for the further development of interventions for those men who do not enter the programme. Much of the information can be obtained by the victim-survivors, some by the men, some by practitioners. (On

ethical aspects of evaluating programmes by victim-survivors see Downes, Kelly & Westmarland, 2014.) Regarding the *victim-survivors*, indicators for successful interventions (see below) should be collected at T0. These indicators will be the central outcome measures at T3 and T4.

- T1: „Start“. The comparison of the sample at T0 and T1 will inform about the first wave of drop-offs, the comparison of T1 and T2 respectively T3 will show which men stay on the programme and who leave it. Changes in life situation and information on context variables will be important from T0 to T3, in order to interpret the various drop-offs. Regarding the *men*, the focus at this point of time will be on variables that should be changed by the programme (outcome measures for the treatment), observed at T1 (e.g. behaviour, attitudes, motivation, risk factors etc.). Again, the *victim-survivors* can be asked about the indicators for successful interventions.
- T2: „During“. Repeated measures are recommended during the intervention. The development of the indicators over time can provide important information, both regarding the men and the victim-survivors. Also, process-related questions can be addressed and connected to the outcomes.
- T3: “End”. The comparison of treatment group and control group, at T1 (pre-treatment) and T3 (post-treatment), is the core of studies to evaluate the impact of interventions. Again, the analysis should include at least two perspectives: the perspective of the victim-survivors, concerning the indicators of success, and the perspective of the men on the programmes respectively the control group, regarding the outcome measures for the treatment itself. Information by practitioners and, additionally, official records can be collected at T3.
- T4: “Follow-up”. The stability of the changes is assessed by a follow-up, ideally six months after the men’s completion of the programme. Again, victim-survivors and men who have been on programmes should be asked, practitioners and official records can serve as additional sources of information.

A drop-off-analysis can be performed for each point of time, for the men as well as for the victim-survivors, if the relevant information (e.g. socio-demographic variables, risk, needs etc.) has been collected at T0.

(b) Designs

Controlled Randomized Trials (RCT) are seen as the gold standard in clinical and psychological research. RCTs rely on the random assignment of research units (e.g. men) to treatment conditions (e.g. intervention vs. no intervention). There are a lot of problems connected to this procedure (e.g. ethical problems), and where RCTs have been realized, they face specific criticism (see Gondolf, 2002, 2012). Even in the criminal justice system, it is often difficult to realize RCTs (see e.g. Dobash et al., 2000), even though probably the best conditions for RCTs are found in jails (see below).

Figure 3. Design of a Randomized Control Design (RCT)

Random assignment of individuals to groups		T1 (pre)	(T2, during TM)	T3 (post)	(T4, follow-up)
	Treatment group	O	X	O	O
	Control group	O		O	O

As a consequence, in various studies the drop-offs, non-attenders or men who have been excluded from programmes have been defined as control groups, whereas completers of DVPPs have been defined as the treatment group. Consequently, treatment group and control group are not equivalent, which constitutes a threat to the internal validity (Shadish, Cook & Campbell, 2002) of such studies. In his famous evaluation study, Gondolf (2002) has found small effects for completers, compared to drop-offs. In order to treat the problem of the non-equivalent control group, Gondolf used propensity score analysis and instrumental variable analysis to adjust the results. These procedures lead to more effectiveness. The principle of such adjustment procedures is the following: The difference between the outcomes of the treatment group and the non-equivalent control group is corrected statistically, by using the differences between the groups (that must be known at the beginning of the study) as a kind of control variables. As a result, an effect size is obtained that reflects what can be expected as effect, assuming that the two groups were as similar as possible. This corrected effect size is seen as more appropriate than the initial uncorrected one by experts in the statistical field (and the uncorrected effect size is seen as inappropriate).

Many previous evaluation studies have focussed on the men on the programmes. Where victim-survivors were included as source of information about the programme's impact, researchers have often been in contact with the perpetrators first, and then have tried to get in contact with the respective victim-survivors. This approach can also be followed in the future; as Gondolf (2002) has demonstrated in his study, up to 60% of the initial victim-survivors could be contacted for the follow-up.

However, there are also alternative approaches, like the following, which has been inspired by the MIRABAL-project⁴. The focus of research can be shifted from the men on the programmes to the victim-survivors in victim support centres. Given the situation that there are geographical regions with victim support centres – but with and without DVPPs, the additional effect of a DVPP can be studied when the victim-survivors' outcomes in such a region or city are compared to a region/city *without* a DVPP (victim support centres are given in both kinds of regions). In this design, the victim-survivors are supported and followed over time (T1-T4), as are their partners – who are referred to a DVPP where there is one available (treatment group), or who cannot be referred to a DVPP because there is none (control group). On the basis of the victim-survivors' information about their partners, it is even possible to perform a kind of matching of treatment group and control group at T0.

⁴ The results of the MIRABAL-project will be presented in early 2015. The project's approach and methodology should be taken up by a future WWP-EN evaluation task force. <https://www.dur.ac.uk/criva/projectmirabal/>

Although it seems useful and meaningful to include the men's perspective in the evaluation and to proceed as outlined above ("Observations / points of time"), we could even think of taking only the victim-survivors' perspective into account, plus some additional information by the practitioners about dropping off, "dose" or completion of programmes. In this case, the DVPP is considered a "black box", administered to a group of perpetrators in a certain dose (depending on their duration of attendance). The treatment group (victim-survivors whose partners are on a programme) is then compared to the control group (victim-survivors whose partners are not on a programme). As a pre-condition, the contexts in the two regions should be very similar, so that differences between treatment-group and control-group can be attributed to the non-/existence of a DVPP. Although the detailed procedures are more complex⁵, the basic idea of studying a DVPP's impact by the differential outcomes of victim-survivors who are on victim support programmes in different regions (with/without DVPPs) is promising.

In a similar manner, also other designs can be developed, e.g. interrupted time-series designs, where victim-survivors are followed over a period of time that corresponds to T0-T5 for a long time, and at a certain moment a DVPP is installed in the region. The hypothesis is that the outcomes which are related to the life situation of victim-survivors improve once the male partners are referred to a DVPP on a systematic basis. Time-series designs are also suitable for single-case research. Especially in settings with small samples, this possibility should not be underestimated. Single-case methodology, also on the statistical level, is a promising development and can be used in the context of evaluation of DVPPs. It is tempting to consider how effects of DVPPs can be studied, based on single-case data. This approach is especially useful when different persons receive different treatments and different treatment intensities ("doses"), depending on their individual needs.

The MIRABAL project has announced to bring another progressive quantitative approach to the evaluation of perpetrator programs, namely *survival analysis*. This kind of analysis originally comes from medical research and the computation of life expectancies, but it is also used in a variety of other fields. Survival analysis starts at mortality tables from which cumulative probabilities for survival are computed. These probabilities, typically displayed as graphs over time, can be given for various subgroups, e.g. persons receiving different treatments. Also, the effects of covariates can be observed. We recommend taking up this innovative approach in a future WWP-EN evaluation task force.

The results of basic evaluation studies as described here will be statements about the additional effect of a DVPP in a CCR (compared to a "CCR without a DVPP").

Irrespective of the design that is realized, different studies can only be compared if the T0-T4-scheme is used in a consistent manner (to know what kinds of men are getting on the respective programme, who is dropping off, and when), and if the same outcome measures are used.

(c) Criteria of success and outcome measures

⁵ For example, groups must be defined post-hoc: Drop-offs on the side of victim-survivors as well as perpetrators; completers and (ex-)partners of completers; the same with excluded and non-admitted; etc.

In a preceding research project for the MIRABAL project, Westmarland, Kelly and Chalder-Mills (2010) have proposed the following “criteria for success” of DVPPs:

- Improved relationship between men on programmes and ex-/partners; respect and improved communication
- Expanded space for action for ex-/partners, restoring their voice, ability to make choices, improved well-being
- Safety and freedom from violence and abuse of women and children
- Safe, positive and shared parenting
- Men’s enhanced awareness of self and others, understanding of the impact of violence on partners and children
- Safer, healthier childhoods, children feel heard and cared about.

These criteria were used to formulate the concrete questions and items for the MIRABAL study (see Westmarland & Kelly, 2012). The perspectives of victims-survivors and children are the primary source of information in this approach to evaluate the impact of a DVPP. Ideally, this information is complemented by the perspective of the men on the programmes, with the outcome measures outlined above. Practitioners and official records can provide additional information which can be used to answer questions that go beyond an outcome evaluation of DVPPs. To anchor future evaluation studies in a good-practice-example, we recommend that a future WWP-EN evaluation task force takes these indicators into consideration.

4.5.2. Multi-site multi-context studies

In the *basic single-context studies*, the focus was on determining the impact of DVPPs, when the conditions are fairly the same in two areas. Now we want to turn to another case, when *similar* DVPPs are operating under *different conditions*. A framework how to think about this constellation is given in figure 4. The context of a DVPP is constituted by the relevant local components that belong to the intervention system in case of domestic violence (police, courts, victim protection etc.), as well as by conditions on a higher level, e.g. the funding situation on the regional level, or laws, campaigns, and preventative measures on a national level. Also international and global conditions can influence the system’s outcomes.

In the *basic single-context studies*, we have assumed that the context is constant, and there are two areas (with and without DVPP) in one country, with the same or a very similar local and regional context. In the *multi-site multi-context studies*, we assume that very similar DVPPs are operated under different conditions, as is to be expected in different countries. The results of Workstream 1 have shown that similar programmes exist in different countries – probably due to the fact that many programmes have developed from the same roots such as the Duluth model or the CHANGE curriculum. When in two or more countries very similar DVPPs can be identified, ideally with the same way of referral (court mandated, self-referred or both), the outcomes of the interventions of the entire systems can be compared. In this case, we can assume that the way the DVPP works is constant, and the context is different, leading to different results. In such a design, the *role of context* and its *impact on the*

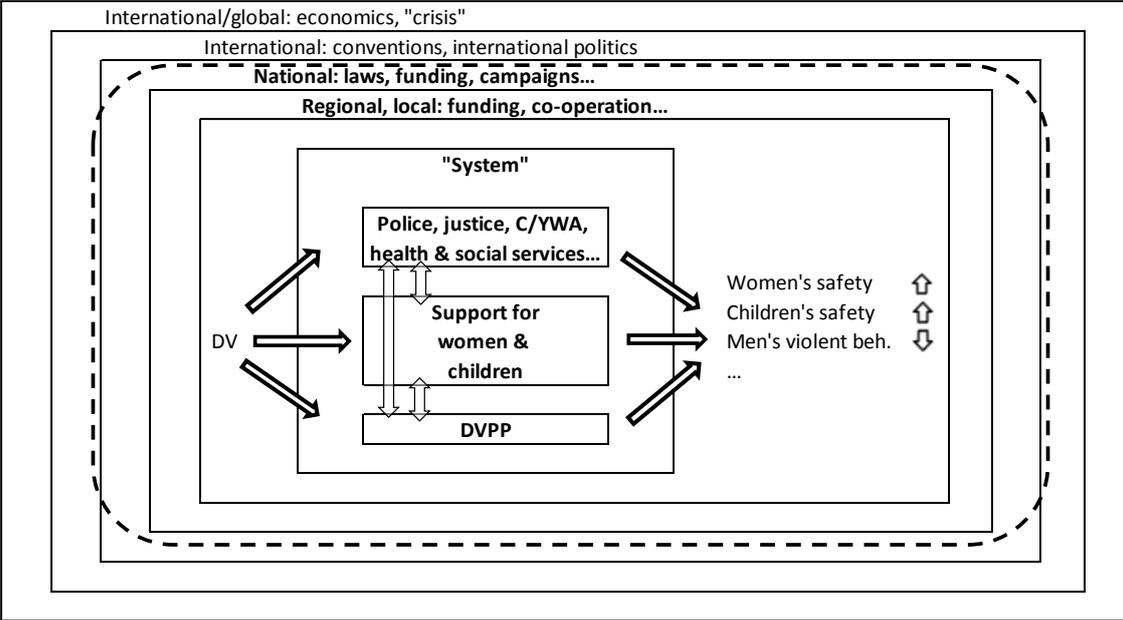
outcomes of a DVPP are compared and studied. It is exactly this question where a multi-country-approach makes sense.

Again, in order to be able to link such a study to other evaluations, it is necessary to apply the same core methodology in all studies. As a result of a multi-site multi-context study, we can expect statements about the influence of context on outcomes of a certain kind of DVPP.

If the context should be analysed further and not be simply treated as a “black box” (with a known overall impact, but with unknown impact of the single components), the challenge is to operationalize the context, in order to determine the influence of the components on the outcomes of the DVPP. Starting from figure 4, the task is to transform the context into variables for a quantitative analysis. Various components need to be “translated” in a way that allows their integration into quantitative statistical procedures:

- On the local level: Support for women and children; the work of the police, criminal justice system, child and youth welfare systems, health and social services;
- On the regional and national level: legal measures, measures regarding violence prevention, funding situation; general gender equality situation.

Figure 4. A framework for different kinds of evaluation studies of DVPPs.



Note. DV... Cases of domestic violence; DVPP... Domestic violence perpetrator programme; C/YWA... Child/youth welfare authority. Cases of domestic violence meet an intervention system that consists of various interacting components. The outcomes of the system’s interventions depend on each component and their interactions, as well as on regional, national and international conditions. From the perspective of a DVPP, the other components of the system and the conditions on the higher levels (national, international) constitute the context of the DVPP.

Such an operationalization is missing today, and its development will require significant resources. Here, we only want to give some hints on how such an operationalization could be done, as a preparation for future *multi-site multi-context studies*.

On the *system level*, the reactions of courts have been identified as relevant influences on the DVPPs' outcomes, in case of court-mandated perpetrators (e.g. Gondolf, 2002). To operationalize this influence, the perspective of men on the programme, practitioners, victims-survivors or experts can be taken as a starting point. If we use the perspective of the *court-referred men on the programme, rated by the practitioners*, the following indicators could be used:

A. Clients' experiences and expectations of court reactions

- "In case of non-compliance, there has been (or can be expected) a swift and consequent reaction by the court." (Value=1)
- "In case of non-compliance, there has been (or can be expected) a slow or erratic reaction by the court." (0,5)⁶
- "In case of non-compliance, no reaction by the court has taken place (or can be expected)." (0)

B. Subjective meaning of the announced or imposed consequences

- "Consequences by the court are meaningful / severe in the eyes of the client." (1)
- "Consequences by the court have some negative meaning for the client." (0,5)
- "Consequences by the court are meaningless in the eyes of the client." (0)

The information per case about A, B, A*B is needed, for all relevant components of a system, and from various perspectives (men, practitioners, victims-survivors). In this way, a lot of case-wise information on context can be entered in a data matrix that is used to perform e.g. a multi-level analysis to assess the influence of the variables and levels on the outcomes of the DVPP. On a higher level, experts can assess the reactions of courts *in general*, which is again entered as a variable into the data matrix. The relevant influences will be determined empirically in such a study.

To determine the relevant components, the development of a tool for comparing contexts can start from comparative research on legal and support measures, such as EIGE (2012). There have also been previous projects which have set out to map and compare contexts regarding measures for victims-survivors in the different European countries (e.g. EUCPN, 2013). This work shows which legal and practical measures are in place, across nations.

Another example for a potentially useful indicator is the Gender Equality Index (GIA), developed by EIGE (2012). Although violence is not part of the GIA due to the lack of comparable data, the GIA can be used as an overall index for the position of a country regarding gender equality, and it can enter an equation as an independent or control variable.

Another interesting approach has been proposed by Hester and Lilley (2014), by pointing to the need to disrupt various risk factors for violence on various levels within an ecological framework (society, institutions, family/peers, individual). This proposal is based on an

⁶ Or a different set of alternatives between 0 and 1.

interactive model of risk factors that has been developed by Hagemann-White and colleagues (see European Commission, 2010). Their model can be used to define pathways for prevention of violence in partnerships; the relevant preventative components can be operationalized and used as variables in multi-level-analyses.

4.5.3. Multi-site single-context studies

The focus of these kinds of studies is to compare different kinds of DVPPs within one context. It is difficult to determine what leads to different outcomes when *different* programmes are studied in *different* contexts – thus it is recommended to hold the context as constant as possible, e.g. by comparing two kinds of programmes in one region.

This approach does not address different ways of working with perpetrators (e.g. more CBT-style or more psychodynamic; or with/without additional services, such as alcohol treatment or psychotherapy), but also different referral pathways could be compared, e.g. CBT-programmes in one region, where one programme works with court-referred clients, the other with self-referred clients etc. The focus should not simply be to determine which programme works better than another, but which approach works better for which clients – which means that it is necessary to study the men's characteristics at intake. The discussion about different treatments for different kinds of perpetrators can be addressed by studies which compare various programmes (but within one context), and also questions of treatment intensity can be dealt with. Standard programmes can provide the benchmark against which programmes with new elements are compared.

Again, it is essential to use the same methodology in all studies, and in this example the information gathered at T0 about the men on the programmes will be highly relevant. The multi-site single-context studies are the right approach to learn more about the interaction of settings, methods, types of perpetrators etc. As results, we can expect conclusions about differential effects of various kinds of DVPPs for various kinds of perpetrators.

4.5.4. Laboratory-like studies

Finally, there is a kind of evaluation studies where it is not meaningful or not possible (e.g. by law) to include the victim-survivors' perspectives (e.g. in the context of perpetrator programmes in jails in Spain). On the other hand, the only RCTs were found in closed institutions like jails and psychiatric hospitals, within the search in Workstream 2 (Hester et al., 2014). Jails provide a setting where it is not possible to observe impacts of perpetrator programmes right away because inmates are not living together with their partners and have only limited contact to them in general. Nevertheless, as part of offender rehabilitation, programmes in jails prepare the participants for their life after imprisonment and have to be seen as an important and meaningful intervention.

In terms of evaluation of DVPPs, the role of programmes in jails can be compared to a laboratory setting in psychological experimentation. In jails, RTCs are possible, in principal. Especially waiting-list designs (i.e. the persons who don't receive treatment and who are

defined as control group get the treatment later on) are a feasible way of realizing RCTs in prisons without major ethical implications. Therefore, studies in jails can aim at improving programmes, by focussing on various treatment components and by studying them in a systematic manner. Effective components can be implemented in programmes in the field (and also be tested and validated by means of *multi-site single-context studies*).

As a result of laboratory-like studies, differentiated statements about effects of treatments in a jail setting (mostly based on only the man's perspective) can be expected, as well as statements about the effectiveness of certain treatment components. Correlation studies with recidivism can be performed – however, recidivism relies on many conditions and influences, not on a man's learning outcome in a programme in jail, so that the expectations concerning this kind of validation should not be too high.

5. Summary and next steps

The initial idea of a big multi-country evaluation study that makes use of the European diversity in approaches towards work with perpetrators, and that answers a range of questions in one go, had to be modified. We focussed on work with perpetrators of domestic violence and on co-ordinated approaches, for various reasons, but also because experts from research and practice advised us to do so in the workshops within this project. We have outlined the basics of a methodology for this field, which still has to be developed in detail, in order to move towards more systematic and harmonized procedures in the evaluation of DVPPs, both on a national and international level. The new international organisation WWP-EN can and should play an important role in this respect because it can develop and promote standards for evaluation of DVPPs among its members, and many future evaluation projects will need to co-operate with these members. A step of high importance that has to be taken now is to designate a group of interested researchers and practitioners within the WWP-EN that feels responsible for these tasks. When they develop methods and standards, they can start from the results of the IMPACT project in general and the recommendations of Workstream 3 in particular.

There are some specifics regarding the evaluation of DVPPs – one of the most important ones is the insight that the most reliable information about relevant change comes from the partners and children of the perpetrators. As a consequence, the co-operation with victim protection organisations is needed, on the concrete level of single centres as well as on the networking level with umbrella organisations like WAVE. It has turned out that a system approach is the most appropriate, which means that we should not strive to evaluate isolated programmes for perpetrators – but to address the whole intervention system. Ways on how we can proceed in this respect have been outlined in this working paper, taking up new developments and good-practice examples of evaluation of DVPPs. We have proposed a framework for evaluation studies that allows to link and compare single studies, and to bring various designs into a meaningful system, based on the developments of the other Workstreams of the IMPACT-project, former studies and some promising developments in the evaluation of DVPPs. Hopefully, this proposal will contribute to overcoming the scattered and rather problematic situation concerning comparability of evaluations of DVPPs today. Step by step, we should become able to detect and demonstrate relevant effect sizes of optimized interventions (“Who needs what”), both for perpetrators and victim-survivors. To have reliable effect sizes for DVPPs are also the basis for meaningful cost-benefit and cost-

effectiveness analyses; in the social and health sectors, such analyses normally prove that the resources allocated for interventions are not only a matter of ethics – but that they are also helping to save public resources. It is a frequent result that the costs of the interventions are low, compared to the costs of the consequences of inactivity.

Further important aspects and considerations will be published in separate papers, based on the work undertaken in Workstream 3 of the IMPACT project.

6. Literature

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

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

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

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

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

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

Council of Europe (2011). Council of Europe Convention on preventing and combating violence against women and domestic violence. Retrieved from <http://www.coe.int/t/dghl/standardsetting/convention-violence/convention/Convention%2010%20English.pdf>

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

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

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

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

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

- EIGE (2013). Gender Equality Index. Report. Vilnius: European Institute for Gender Equality. Retrieved from <http://eige.europa.eu/content/document/gender-equality-index-report>
- EUCPN (2013). Tackling domestic violence in the EU – policies & practices. In EUCPN Secretariat (eds.), EUCPN Toolbox Series, no. 4, European Crime Prevention Network: Brussels.
- European Commission (2010). Feasibility study to assess the possibilities, opportunities and needs to standardize national legislation on violence against women, violence against children and sexual orientation violence. Luxembourg: Publications Office of the European Union.
- Geldschläger, H., Ginés, O., Nax, D. & Ponce, A. (2014). Outcome measurement in European perpetrator programmes: A survey. Retrieved from http://wwp.webseiten.cc/fileadmin/WWP_Network/redakteure/IMPACT/Daphne%20III%20Impact%20Working%20Paper%201%20rev.pdf
- Gondolf, E. W. (2002). Batterer intervention systems: Issues, outcomes, and recommendations. Thousand Oaks, CA: Sage.
- Gottzen, L. (2013). Encountering violent men: Strange and familiar. In B. Pini & B. Pease (eds.), Men, masculinities and methodologies. New York: Palgrave Macmillan.
- Hamilton, L., Koehler, J. A. & Lösel, F. (2013). Domestic violence perpetrator programs in Europe, Part I: A survey of current Practice. *International Journal of Offender Therapy and Comparative Criminology*, 57 (10), 1189-1205.
- Hester, M. & Lilley, S.-J. (2014). Preventing violence against women: Article 12 of the Istanbul Convention. A collection of papers on the Council of Europe Convention on preventing and combating violence against women and domestic violence. Strasbourg: Council of Europe.
- Hester, M., Lilley, S.-J., O'Prey, L. & Budde, J. (2014). Overview and analysis of research studies evaluating European perpetrator programmes. Retrieved from http://wwp.webseiten.cc/fileadmin/WWP_Network/redakteure/IMPACT/Daphne%20III%20IMPACT%20Working%20Paper%202%20Overview%20and%20analysis%20of%20research%20studies%20evaluating%20European%20perpetrator%20programmes.pdf
- Johnson, H. (1998). Rethinking survey research on violence against women. In R. E. Dobash & R. P. Dobash (eds.), Rethinking violence against women. Thousand Oaks, CA: Sage.
- Kutrovátz, K. (2013) Work with Perpetrator – European Network national report Hungary. Retrieved from http://wwp.webseiten.cc/fileadmin/WWP_Network/redakteure/Reports_2013/National_report_Hungary.pdf
- Patulny, R., Pini, B. (2013) Counting men: quantitative approaches to the study of men and masculinities. In B. Pini & B. Pease (eds.), Men, masculinities and methodologies. New York: Palgrave Macmillan.
- Payarola, M. A. (2011) Description of a voluntary attendance program for abusive men. In C. Blazina & D. S. Shen-Miller (eds.), An international psychology of men. Theoretical advances, case studies, and clinical innovations. New York: Routledge.

- Phillips, R., Kelly, L. & Westmarland, N. (2013) Domestic violence perpetrator programmes: A historical overview. London and Durham: London Metropolitan University and Durham University.
- Price, B. J. & Rosenbaum, A. (2009). Batterer intervention programs: A report from the field. *Violence and Victims*, 24 (6),747-770.
- Rossi, P. H., Lipsey, M. W. & Freeman, H. E. (2004). *Evaluation. A systematic approach* (7th ed.). Thousand Oaks: Sage.
- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.
- Silverman, D. (2013). *Prowadzenie badań jakościowych*. Warszawa: Wydawnictwo Naukowe PWN.
- Stufflebeam, D. L. & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco: Jossey-Bass.
- Westmarland, N. & Kelly, L. (2012) (with Kelly). Why extending measurements of 'success' in domestic violence perpetrator programmes matters for Social Work. *British Journal of Social Work*, 43(6), 1092-1110.
- Westmarland, N., Kelly, L. & Chalder-Mills, J. (2010). *What Counts as Success?* London: Respect.
- World Health Organization (2002). *World report on violence and health. Summary*. Retrieved from http://www.who.int/violence_injury_prevention/violence/world_report/en/summary_en.pdf
- WWP European Network (2008). *Guidelines to develop standards for programmes working with male perpetrators of domestic violence*. Retrieved from <http://www.work-with-perpetrators.eu/index.php?id=86>