



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

Overview and Analysis of Research Studies Evaluating European Perpetrator Programmes

Working paper 2 from the Daphne III project “IMPACT: Evaluation of European Perpetrator Programmes”

Marianne Hester, Sarah - Jane Lilley, Llorenc O’Prey & Jeppe Budde



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

Index

1. Introduction	2
2. Method	3
2.1 Identification of studies	4
2.2 Study selection	6
2.3 Data extraction	12
2.4 Quality assessment	14
2.5 Data synthesis	14
3. Results	15
3.1 Evaluation design	15
3.2 Sample profile	16
3.3 Outcome measures	25
3.4 Follow-up periods	33
3.5 Impact	33
3.6 Limitations	35
3.7 Standard of reporting	35
4. Summary and conclusion	35

1. Introduction

Evaluating whether domestic violence perpetrator programmes contribute to the safety of women and children victims/ survivors is essential for both policy makers and for practitioners¹. However, until very recently there was no synthesised body of European evidence on ‘what works’ in domestic violence perpetrator programmes and issues of transferability means that existing evidence (e.g. from North American studies) cannot easily be generalised to a European context². In order to fill the existing knowledge gap about evaluations of perpetrator programmes across Europe, the Daphne III IMPACT project aimed to:

- Provide an overview of outcome monitoring practices within perpetrator programmes across Europe (workstream 1)
- Provide an overview of the research /evaluations of these programmes (workstream 2)
- Identify the possibilities and challenges of a multi-country, European-wide evaluation methodology (workstream 3)
- Develop a monitoring/evaluation toolkit that can be used by perpetrator programmes in future (workstream 4).

This working paper describes the methodology and results of the work undertaken in workstream 2. The main objective of workstream 2 was to provide detailed analysis of a range of evaluation research studies linked to European perpetrator programmes, in order to provide criteria for robust evaluations and to feed into the development of a monitoring/evaluation toolkit [workstream 4]. The focus was not on the day-to-day outcome monitoring practices of perpetrator programmes but on the scientific process and outcome research.

Specifically, workstream 2 aimed to develop

- detailed knowledge about the approaches used in evaluation research studies across Europe, with particular emphasis on the methods, input, output and measures of outcome
- a set of criteria related to scientific robustness that can accommodate realistic approaches and a variety of methods and thus point to a “new generation” of evaluation research.

¹ Geldschläger, H et al (2014) *Outcome measurement in European perpetrator programmes: a survey*. Working paper 1 from the Daphne III IMPACT project. Barcelona, January 2014.

² Akoensi, TD., Koehler, JA., Lösel, F., and Humphreys, DK (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) pp 1206-1225

The following questions underpinned the work:

1. What types of evaluation are being used across Europe?
2. In particular what methods, designs, and outcome measurements are being used?
3. Are the outcome measures being used sensitive to different groups of perpetrators or the potential outcomes of the programmes?
4. Are there differences or similarities between (and within) different regions of Europe? (Central, East, North, South, West)
5. What are these evaluations showing in terms of the impact of perpetrator programmes across Europe?

This working paper describes the methodology used in the overview, the results of the analysis of the European evaluations of perpetrator programmes, followed by discussion and some conclusions. More detailed analysis of the evaluation studies identified and included in the overview will be provided in two separate scientific articles.

The partners of the IMPACT project are:

- Dissens – Institut für Bildung und Forschung e.V., Berlin, Germany
- Askovgaarden, Copenhagen, Denmark
- CONEXUS, Barcelona, Spain
- Men’s Counselling Centre Graz, Austria
- RESPECT, London, UK
- University of Bristol, UK
- WAVE, Vienna, Austria

The authors would like to thank all the partners of the IMPACT project, and in particular David Nax (Dissens), Heinrich Geldschläger (CONEXUS) and Christian Scambor (Men’s Counselling Centre Graz) for their help and support with the workstream 2 data extraction process.

2. Method

The aim of the overview was to assess the scope of existing European research studies, the range of evaluation designs and outcome measures used, and the results emerging from these studies. It was envisaged that examination of the strengths and limitations of such a wide range of studies would inform the criteria for robust evaluation of European perpetrator programmes, feeding into the IMPACT monitoring/evaluation toolkit (workstream 4).

A recent systematic review of European evidence on the effectiveness of domestic violence perpetrator programmes³ searched through over 10,000 titles and discovered only twelve studies that evaluated the effectiveness of a perpetrator programme in some systematic manner. The review concluded that while the evaluations showed various positive effects after intervention the methodological quality was insufficient to develop strong conclusions or estimate an effect size, suggesting that evaluation of domestic violence perpetrator programmes in Europe must be improved. We have taken a wider approach and attempted to look at all evaluations of perpetrator intervention⁴ in Europe (published and unpublished) in order to gain a better understanding of the range and variety of evaluation approaches and design across Europe.

Due to the inclusion of studies written in any EU/accession language the studies were divided into five regions of Europe (Central, Eastern, Northern, Southern and Western) and searches, translation and data extraction conducted by different members of the international IMPACT project team situated in different countries across Europe. The research study reports were uploaded and stored on a dedicated web-based workspace (and allocated a unique reference ID) for ease of access and this was used to manage the process.

2.1 Identification of studies

As we wanted to access both published and unpublished evaluations, a variety of search methods and comprehensive approach to sources were employed by the project team. The evaluation studies were therefore identified via:

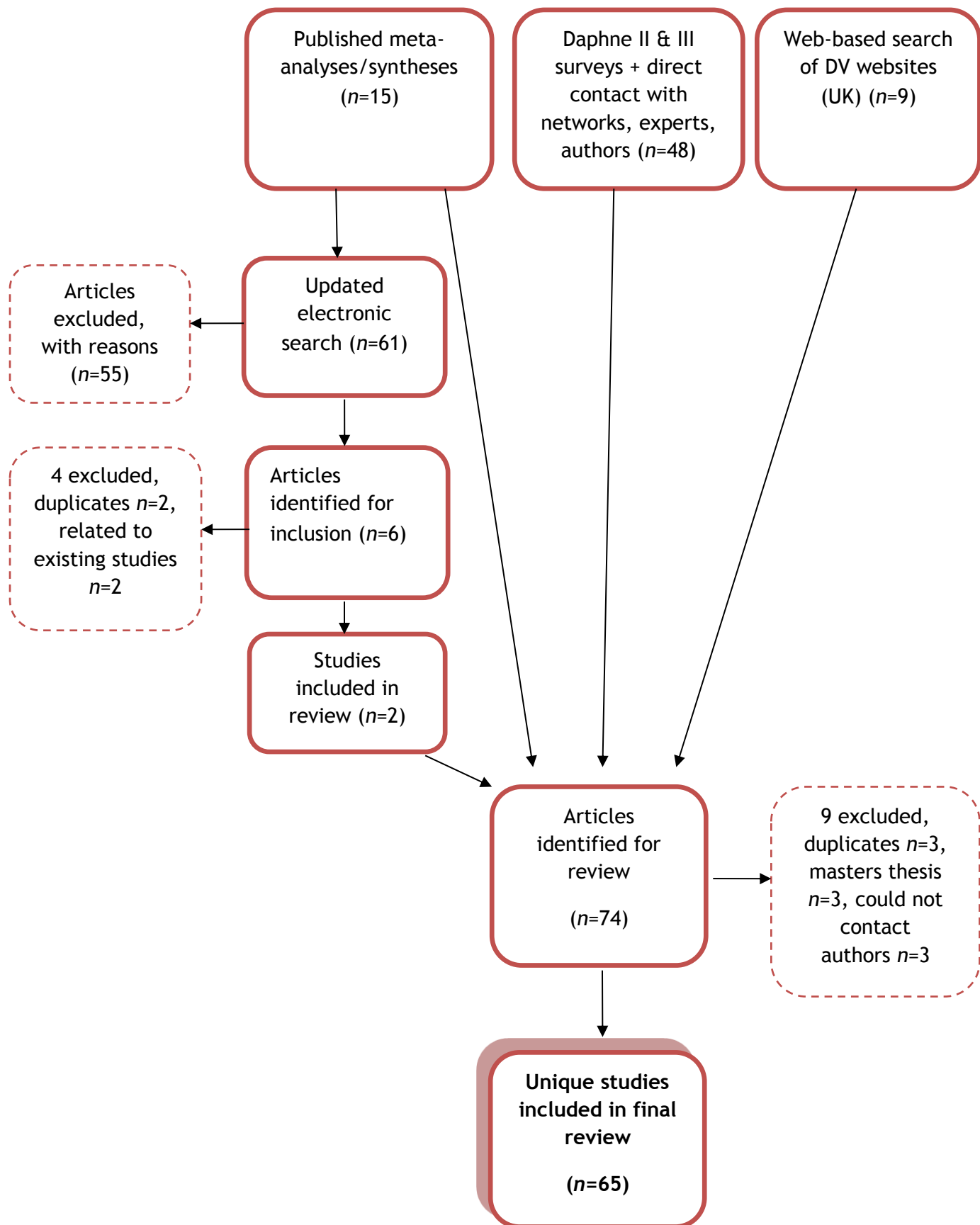
- Existing (published) meta-analyses/syntheses which included European studies (Feder, Hester et al 2008; Arias et al 2013; Akoensi et al 2013 and NICE 2014). A total of fifteen studies were identified from these publications, originating from the UK ($n=7$), Spain ($n=5$), Sweden ($n=1$), Finland ($n=1$) and Germany ($n=1$).
- An update of the systematic search of electronic databases included in the systematic review conducted by Akoensi et al (2013). As it was unclear precisely when the original search took place it was assumed that it was completed sometime in 2012. A time filter was therefore placed on all electronic searches before January 1st 2012 to ensure complete coverage.
- The two separate Daphne II & III Work With Perpetrators Surveys of perpetrator programmes in EU/accession countries in 2007/8 and 2013⁵.
- Direct contact with European perpetrator programme network/s; study authors and experts.
- Additional searches of specialist domestic violence websites (UK).

³ Ibid, p 3.

⁴ For the purposes of consistency of terminology throughout this document we will use the word 'intervention' to refer to perpetrator programmes or 'treatment'.

⁵ <http://archive.work-with-perpetrators.eu/en/resources.php>

Flowchart of study identification process



2.2 Study selection

In order to overcome some of the issues associated with publication bias, and to capture as many of the European evaluation studies as possible, we employed wider inclusion criteria than those used for the four meta-analyses identified as part of the search (see table 1 for comparison of our IMPACT W2 inclusion criteria and those of the meta-evaluations). Our criteria included:

- Evaluations of programmes working with male domestic violence perpetrators conducted in any European country (EU/ accession countries).
- Studies written in any EU/ accession language and that were either published (i.e. formally issued/controlled by a commercial publisher) or included in the grey literature (i.e. reports not widely distributed or commonly used in abstracts or indexes, for example, reports produced or published by universities or academic research units, Government reports, programme / funder reports and including PhD studies).
- Evaluations that were process/ implementation or outcome focussed (or those that included *both* process and outcome measures).
- Studies that were of experimental design (i.e. studies that used a control group, for example, RCT or other control); quasi-experimental design (i.e. studies that used comparison, for example, of different sites, different cohorts, different interventions, that were pre-post or follow-up); and non-experimental studies (i.e. that used single group or amalgamated data, and were generally descriptive, pre-post, and/ or follow-up).
- Studies using any outcome measures.
- No reporting requirements were specified.

Table 1 Study Inclusion Criteria – comparison of inclusion criteria for published meta-analyses/syntheses and Daphne III IMPACT overview

Criteria	Feder, Hester et al 2008 ⁶ (n=31)	Arias et al 2013 ⁷ (n=19)	Akoensi et al 2013 ⁸ (n=12)	NICE 2014 ⁹ (n=34)	DAPHNE III IMPACT 2014 (n=65)
Type: outcome, process	Outcome	Outcome (recidivism for completers - measured by official reports or couples reports - as proportion of re-offending batterers –other offences excluded – during follow-up period)	Outcome (both attitudinal and behavioural measures)	Outcome and process	Any
Design: RCT, experimental, Pre-post, descriptive	Control studies with a follow-up of at least 6 months	Not specified or clear from the report	Any. Minimum = outcome measures pre & post treatment (Level 2 Maryland Scale of Scientific Methods)	RCT, case control study, interrupted time series, cohort study, cross-sectional study, systematic review or qualitative study	Any
Methods: Standard measures, interviews	Any	Not specified or clear from the report	Any	Any	Any
Location	Any	Not specified or clear from the report	EU and accession	countries in the Organisation for Economic Co-operation and Development (OECD)	EU and accession
Target population	Male DV perpetrators	DV perpetrators (unclear as to whether male perpetrators only)	DV perpetrators (male or female, convicted offenders or voluntary non-offenders).	people who perpetrate domestic violence and abuse	Male DV perpetrators

⁶ Feder, G., Hester, M., Williamson, E and Dunn, E (2008) *Behavioural interventions to reduce intimate partner violence against women*. In J.A Trafton and G.P Williams (eds) Best practices in behavioural management of health from pre-conception to adolescence, Vol III. Institute for Disease Management/ Institute for Brain Potential pp522-567

⁷ Arias, E., Arce, R., and Vilarino, M (2013) *Batterer intervention programmes: A meta-analytic review of effectiveness*. *Psychosocial Intervention* 22 (2013) 153-160

⁸ Akoensi et al (2013) *op. cit.*

⁹ National Institute of Clinical Excellence (NICE) (2014) *Domestic violence and abuse: how services can respond effectively*. <http://publications.nice.org.uk/lgb20>

Intervention	Perpetrator programmes, Anger management, Couples counselling	Duluth, CBT or OTI programmes	Any programme designed to alter the attitudes and/or behaviours of DV perpetrators	any intervention or approach to identify, prevent, reduce or respond to domestic violence and abuse between adults and young people who were, or had been, intimate partners & focused on healthcare, social care or specialised services that deal with domestic violence and abuse	Any
Language	English	Spanish & English	Any European	English	Any European
Publication	Published	Any	Both published & unpublished	Both published & unpublished	Any
Database / searches	Medline; Social Science Citations; Institute of Social Sciences; Sage journals; Taylor & Francis journals;	Broad spectrum of databases (small and specialised with quality control e.g. Scopus & Web of Knowledge); gender violence observatories; researchers in field (authors of retrieved & excluded articles); reviewing & cross-referencing reference sections of previous meta-analysis	On-line computerised databases & specialist journal archives; meta-analytic & systematic reviews dealing with DV PPs; academics & experts contacted; respondents to Daphne II WWP Survey contacted individually.	A number of databases were searched in May 2012 for randomised controlled trials (RCT), case-control studies, interrupted time series, cohort studies, cross-sectional studies, observational studies, systematic reviews and qualitative studies.	Identification was via existing meta-analyses/syntheses, two rounds of EU/accession countries surveys of perpetrator programmes in IMPACT and predecessor (2007/8 & 2013), plus direct contact with European perpetrator programme network/s, plus further internet searches using DV sites.

Reporting requirements	None specified	Sample size; recidivism rate for treatment completers; recidivism measured by official reports and couple reports; description of treatment theoretical approach, contents & duration; recidivism during follow-up period (over 6 months)	None specified	None specified	None specified
-------------------------------	----------------	---	----------------	----------------	----------------

Updated electronic search

The updated electronic searches used the same search strings as the Akoensi et al (2013) meta-analysis, as follows:

Domestic violence or Domestic assault or Batterer or Family violence or Physical abuse or Spousal abuse or Interfamily violence or Intimate partner violence or Duluth

And

Program* or Treat* or Intervention* or Therapy Counsel* or Rehab* or Court decisions or Mandated court decisions or prison*

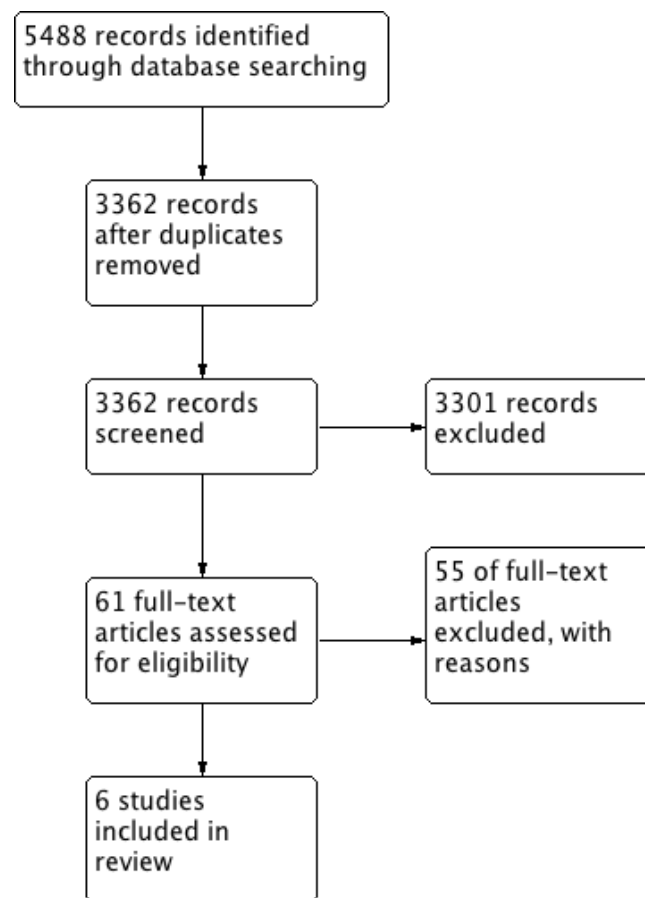
And

Effect* or Outcome* or Eval* or Experiment* or Randomi*ed controlled trials or Quasi experiment* or Trial or Empirical or Recidiv*

The following databases were searched using the search strings:

- ♦ International Bibliography of the Social Sciences (IBSS)
- ♦ PsycInfo
- ♦ PubMed
- ♦ EmBase
- ♦ ISI Web of Knowledge

Flowchart of Updated Electronic Searches and Screenings



Inclusion criteria for updated electronic search

The same regional, methodological, population and intervention criteria were applied. The search produced 6 unique studies that were deemed to fall within the inclusion criteria, three of which had already been identified in our previous search strategy and included in the overview / dataset.

DAPHNE II and III survey sample

The programmes who reported in the DAPHNE II survey that they had had an evaluation were followed up on a regional basis to obtain the evaluation studies. Some were not actual evaluations but monitoring reports. This process was updated when the DAPHNE III survey became available. In addition to the surveys, direct contact was made with European perpetrator programme network/s; individual study authors and experts in the field. Altogether this yielded 48 unique studies for inclusion.

Further UK-based web search

A further web-based search of specialist domestic violence websites in the UK was conducted to identify any additional evaluation study reports / articles. This yielded a total of 9 reports for inclusion in the database.

2.3 Data extraction across all data

The database of European perpetrator programme evaluation studies was compiled in English. At the beginning of the process the other European language reports were summarised in English and data was extracted using a specifically designed template (to capture article reference and region; background and context of programme and evaluation; research method for evaluation; method of intervention and output; and the evaluation findings). A more detailed data extraction spreadsheet was then developed (using Microsoft Excel) in order to combine data extraction and data entry and better facilitate data analysis.

Where multiple reports relating to the same evaluation were identified the individual related references were recorded, but the data was combined and counted as a 'unique' study for analysis purposes.

Three main categories of data were extracted by one main assessor and checked by another, with the following information from each evaluation recorded on the spreadsheet:

I. The source and programme:

- Unique study ID
- Full bibliographic reference
- Type of publication (published/ 'grey'/student thesis)
- Language of original study report / article.
- Participation/ paths of entry and intervention setting
- The theoretical model that the programme/intervention is based on
- Programme structure (e.g. rolling or fixed programme; number/frequency/length of sessions, minimum & maximum period of intervention; individual sessions followed by group sessions; size of groups, whether closed / open groups) and
- The context within which the programme is set and run i.e. whether and how it fits with the wider response to domestic violence and abuse (e.g. is it part of an existing multi-agency approach or a stand-alone service?)

II. The evaluation:

- Type of evaluation i.e. *implementation/process* (processes of embedding and/or delivering programme etc); *outcome* (demonstrates effects against targeted outcomes); *impact* (broader effects intended or unintended of programme); *cost-benefit* or *cost-effectiveness* or whether a combination of these types

- Focus of the evaluation i.e. what exactly is being evaluated and why? (may be whole or part of a programme/intervention) based on aims and objectives / hypotheses / research questions etc
- Method. A summary of the evaluation methods employed
- Evaluation design i.e. experimental, quasi-experimental or non-experimental
- Total sample size and drop-out rate (between the start of the evaluation and the end and/or follow-up)
- Outcome measures i.e. summarising **what** is being measured (e.g. change in attitude or abusive behaviour of perpetrator (reductions or changes in physical violence, controlling behaviour, emotional, sexual or other types of abuse? Increase in self-awareness or empathy? Increase in safety for women partners? Increased resilience to, or a reduction in, repeat victimisation? Increase in self-esteem of women partners/victims?) **how** it is measured (e.g. through standardised or specifically designed tests/ inventories, questionnaires, interviews, official or project data) and **who** is the source of data? (e.g. male/female perpetrator/participants, women partners/ex-partners or family members, agency staff, programme staff etc)
- Results i.e. a summary of the main evaluation findings
- Limitations of the evaluation design and method (as outlined in the original report plus those based on the assessors' judgment)
- Statistical results. Recorded separately from the overall 'results' field for ease of reference and meta-analysis/synthesis.

III. Shape and size of the sample at the different stages of the evaluation process:

- **Time 0** (T0) Details of the sample size and profile prior to the start of the intervention¹⁰ and /or at the referral stage (e.g. socio-demographics, substance misuse problems, mental health problems of participants', who is excluded or who drops out at this stage?). What measures are taken at this stage?
- **Time 1** (T1) Details of the sample size and profile at the start of intervention. What measures are taken at this stage?
- **Time 2** (T2) Details of the sample size and profile during intervention (including – where available- the number of times/ when measures are used (e.g. T2a at beginning of programme, T2b at exit if during programme etc)
- **Time 3** (T3) Details of the sample size and profile at the end of intervention (including –where available – details of the completers and non-completers etc). What measures are taken at this stage?
- **Time 4** (T4) Details of the sample size and profile at the follow-up stage. What measures are taken at this stage? (and including –where available - follow-up with completers and non-completers, number of times/when measures are taken (e.g. T4a 3 months, T4b 6 months, T4c 12 months)

¹⁰ For the purposes of consistency within this document we will use the word 'intervention' to refer to perpetrator programme / treatment.

To improve validity and reliability the data extraction spreadsheet was further developed and piloted within the team and instructions for completion provided to the assessors. Team meetings (facilitated by telephone / Skype) were held to manage progress and issues arising from the data extraction process.

Dividing the studies into regions allowed different team members (with different backgrounds and levels of involvement in the project) to conduct the data extraction independently. To improve validity and reliability (and reduce the risk of bias due to information being interpreted and extracted differently/subjectively) a second reading of the studies and data extraction form was conducted by different team members to check for consistency and accuracy of assessment and interpretation of the data.

2.4 Quality assessment

For the purposes of this study it was important to record the limitations of each study's design and method on the data extraction spreadsheet along with comments relating to the quality of reporting. Subjective interpretation of the quality (strengths and weaknesses) of each evaluation study can introduce a risk of bias into the data extraction process. The limitations as outlined in the study report itself (e.g. risks of bias and issues of generalizability) were recorded in addition to the individual assessors' judgement on the quality, strengths and weaknesses.

2.5 Data synthesis – interpreting the evidence

A narrative synthesis will be conducted on the 65 evaluation studies identified and, where possible, a synthesis of the statistical results will be conducted quantitatively. These will be published separately.

3. Results

In total our search found 65 unique evaluation studies. Thus we included 50 studies that were not already included in the four published meta-analyses/syntheses. The 65 studies originated from 13 European countries. In order of number of studies the list was: Spain (n=22), UK (n=19), Germany (n=6), Switzerland (n=4), Finland (n=4), Sweden (n=2), Austria (n=2), Ireland (n=1), Denmark (n=1), Iceland (n=1), Croatia (n=1), Netherlands (n=1) and Portugal (n=1).

Table 2 Language and publication

	North (n=8) Sweden, Denmark, Finland, Iceland	South (n=23) ¹¹ Spain, Portugal	West (n=20) UK, Ireland	East (n=1) Croatia	Central (n=13) Germany, Austria, Switzerland, Netherlands
Language	English 2 Other EU 5 <i>unknown = 1</i>	English 4 Other EU 19	English 20 Other EU 0	English 1 Other EU 0	English 1 Other EU 12
Publication	Published 3 Grey lit 2 <i>unknown = 3</i>	Published 12 Grey lit 7 PhD thesis 4	Published 4 Grey lit 16	Published 0 Grey lit 1	Published 10 Grey lit 3

- An equal amount of studies were published (n=29) as were included in the ‘grey’ literature i.e. not widely distributed or commonly used in abstracts or indexes (n=29).

3.1 Evaluation design

Table 3 Study design and method

	North (n=8) Sweden, Denmark, Finland, Iceland	South (n=23) Spain, Portugal	West (n=20) UK, Ireland	East (n=1) Croatia	Central (n=13) Germany, Austria, Switzerland, Netherlands
Type:	outcome 2 process 3 both 2	outcome 19 process 1 both 3	outcome 6 process 1 both 13	outcome 1 process 0 both 0	outcome 6 process 0 both 6

¹¹ Figures changed as of 01.04.14 as 6 reports related to 3 unique studies (SSP1 reports on outcome of study but SSP3 is ex post facto study examining process using the same dataset). Figures in this table assume 23 unique studies in Southern region.

	unknown= 1			unknown =1	
Design	RCT 0	RCT 1	RCT 0	RCT 0	RCT 1
	Oth control 0	Oth control 1	Oth control 0	Oth control 0	Oth control 1
	Quasi-exp 0	Quasi-exp 7 ¹³	Quasi-exp 3	Quasi-exp 1	Quasi-exp 1
	Pre-post ¹² 3	Pre-post 20	Pre-post 11	Pre-post 1	Pre-post 7
	Descriptive 1	Descriptive 0	Descriptive 8	Descriptive 0	Descriptive 5
		unknown= 2			
	(k=4)	(k=22)	(k=19)	(k=1)	(k=12)

- The majority of papers reported study designs that were focused on outcomes ($n=34$, both implementation/process and outcome $n=24$, process $n=5$).
- Most studies were of a non-experimental design ($n=25$, quasi-experimental $n=12$, experimental $n=4$) and there was substantial use of pre-post testing as a means of exploring change across time ($n=42$).
- Two studies employing an experimental RCT design were identified, originating from Spain (SSP20)¹⁴ and the Netherlands (CNE1)¹⁵.
- Two studies (originating from Spain) used trial designs that compared two different modes of delivery to that of a control group (SSP20 and SSP22¹⁶).

The main limitations of the European studies centred around the lack of a control group in many studies, and where control groups are used they involved small samples. These issues can severely limit the potential conclusions that can be drawn from the results about programme effectiveness.

3.2 Sample profile

An important part of the analysis of evaluations involves looking at the 'story' of what happens to the sample over time, and the data extraction process was designed to capture

¹² All studies with a pre-post design i.e. including experimental, quasi-experimental and non-experimental

¹³ 8 of the 9 quasi-exp studies have pre-post design, the remaining study only collected data post intervention

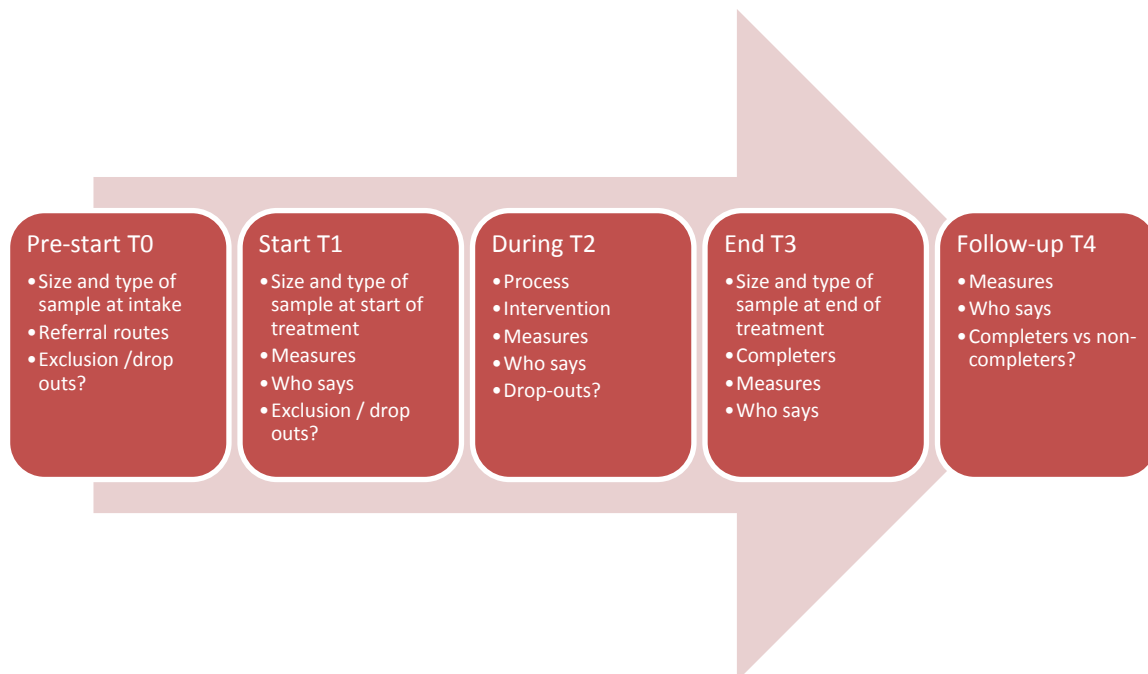
¹⁴ Rodríguez-Espartal N & Lopez-Zafra E (2013) *Programa emocional para presos por violencia de género (PREMOVIGE): Efectividad en variables cognitivas y conductuales [Emotional program for inmates imprisoned for gender violence (PREMOVIGE): Effectiveness in cognitive and behavioral variables]* in Psychosocial Intervention, 22, 115-23.

¹⁵ Kraanen, F., Vedel, E., Scholing, A., & Emmelkamp, P. (2013) *The comparative effectiveness of Integrated treatment for substance abuse and partner violence (I-StoP) and substance abuse treatment alone: a randomised controlled trial* in BMC Psychiatry 2013, 13: 189.

¹⁶ Subirana-Malaret M y Andres-Pueyo A (2013) *Retención proactiva y adherencia terapéutica en programas formativos para hombres maltratadores de la pareja [Proactive retention and therapeutic adherence in programs for male perpetrators of intimate partner violence]* in Psychosocial Intervention, 22: pp95-104.

the shape and size of the sample at the different stages of the evaluation process (see Figure 1 below).

Figure 1



Defining commencement and completion of intervention

- Most programmes subjected to evaluation comprised pre-intervention assessments, followed by individual counselling sessions which had to be completed in order to start the group sessions.
- Engagement or commencement of intervention however was not always explicitly defined in the evaluation reports.
- In some studies the baseline sample was taken from the pre-intervention assessment phase but for most commencement of intervention was defined as completion of the pre-assessment stage and commencement of the individual counselling sessions (sessions varied in number and duration and generally had to be completed before a participant was classed as 'engaged' and able to move onto the 'core' intervention which was the (usually weekly) group work sessions.
- Few studies explicitly defined intervention completion. For most, 'completion' was defined by attendance at either all or a high number of the 'core' group sessions, which usually followed a set number of individual counselling sessions pre-group work.
- Where it was reported the definition of intervention 'completion' also varied widely from programme to programme. For example in one study 'satisfactory programme completion' was defined as the successful achievement of solution-building & safety

goals (WUK7), in another study completers had attended all 21 sessions (CCHDE4) and another study reported that the intervention was 'closed' successfully where there was settlement between victim and perpetrator (CAU5).

- Many studies reported attrition occurring between the individual sessions and the 'core' intervention group sessions.
- Many studies collected outcome data upon completion of intervention, which can lead to reporting of 'false-positive' effect of intervention.

Size and shape of the sample

- Sample sizes varied greatly across studies, with the smallest $n=10$ (SSP23) and the largest $n=770$ (SSP14).
- Over half of the studies examined (57%) reported sample size at intake and completion as a minimum ($n=37$).
- 18 studies (28%) reported about the sample at intake, commencement and completion.
- Only 2 studies reported information about the sample size at every stage i.e. at intake, during intervention, upon completion and at follow-up (SSP1 and WUK15).
- A small number of studies ($n=3$) only collected data post intervention completion.

Information about the sample

- The type of information reported about the sample varied greatly across studies.
- Socio demographic information for participants was mainly measured at intake/ pre-assessment and commencement stage only. Few studies provided comparative data for those that 'completed' the intervention.
- A handful of studies provided information on drop-outs beyond numbers i.e. reasons for dropping out and characteristics / demographics.
- 20 studies provided information on the programme entry pathways. For the Southern studies most samples were drawn from the judicial system and referred to programmes as an alternative to prison in which attendance was compulsory ($n=15$). In some participants could attend voluntary ($n=8$), and some programmes applied a combination of voluntary and court mandated attendance ($n=3$).
- 10 studies reported some level of comparative analysis of completers and non-completers (SSP15, WUK4, WUK7, WUK19, CDE1, CDE2, CDE3, CCHDE1, CAU6, CNE1).

Table 4 Sample size and shape

Sample size	At intake*	Drop-outs	Start* of intervention	End* of intervention	Follow-up*
1-10	SSP23 (n=10)		SSP23 (n=10) WEI2 (n=2)	SSP23 (n=10) ¹⁷	
11-25	WUK2 (n=21) CDE9 (n=20) CCHDE3 (n=22)		WUK2 (n=18) CDE9 (n=16) CCHDE3 (n=15)	WUK2 (n=3) WUK5 (n=18) CDE9 (n=13) CCHDE3 (n=15)	WUK3 (n=20) WUK5 (n=18) CCHDE3 (n=22)
26-50	SSP18 (n=38) SSP20 (n=36) WUK11 (n=32) CDE3 (n=46) CCHDE1 (n=23) CCHDE4 (n=41)	WUK11 (n=17) CCHDE1 (n=3)	WUK1 (n=43) SSP18 (n=38) SSP20 (n=36) SSP24 (n=31) WUK11 (n=15) WUK17 (n=26) CDE3 (n=38) CCDE1 (n=17) CCHDE4 (n=39)	WUK1 (n=21) SSP18 (n=38) ¹⁸ SSP20 (n=36) ¹⁹ SSP24 (n=16) ²⁰ WUK11 (n=5) WUK17 (n=14) CDE3 (n=31) CCHDE1 (n=14) CCHDE4 (n=14 completers)	WUK17 (n=9)
51-75	SPO1 (n=55) SSP15 (n=70) SSP16 (n=62)	SSP15 (n=4) SSP16 (n=4) WUK7 (n=18)	- SSP11 (n=62: 32 intervention & 30 control) SSP15 (n=52) SSP16 (n=47) WUK7 (n=52) ²¹	SPO1 (n=55) SSP11 (n=62) ²² SSP15 (n=48 completers) SSP16 (n=43) WUK7 (n=34)	

¹⁷ Not specified in report but no reported attrition

¹⁸ Not specified in report but no reported attrition

¹⁹ Not specified in report but no reported attrition

²⁰ Not specified in report but report states 48% of 31 men dropped out

²¹ Total sample n=68 but this included 52 men and 16 women perpetrators. Only men counted here.

²² Not specified in report but no reported attrition

	WUK13 (n=70)		WUK13 (n=32)		
	CNE1 (n=52)		WE11 (n=72) CNE1 (n=52)		CNE1 (n=19)
76-100	SSP9 (n=100)				SSP9 (n=40: 32 intervention, 8 non-intervention)
			WUK6 (n=84: 52 intervention & 32 in control)		WUK6 (n=84: 52 intervention & 32 in control)
	WUK10 (n=97)		WUK10 (n=76)		WUK10 (n=47)
101-150	SSP2 (n=148)				SSP2 (n=101)
	SSP4 (n=142)	SSP4 (n=40)	SSP4 (n=102)		
	SSP7/12 (n=109)				SSP7/12 (n=73)
	SSP22 (n=142: proactive groups, non-proactive groups) ²³	SSP22 (50.1% of proactive group dropped out & 63.2% non-proactive group dropped out)			SSP22 (n=?)
	WUK16 (n=115)				WUK16 (n=26)
	CDE7 (n=104)		CDE7 (n=48)		CDE7 (n=48) ²⁴ CDE7 (n=4)
151-200	CDE5 (n=157)	CDE5 (n=56)	CDE5 (n=15) NSE1 (n=188)		CDE5 (n=7) NSE (n=140)

²³ Size and composition of the two groups not reported

²⁴ Of the 104 on the programme, data for 48 was collected - however not clear whether these were completers or not?

201-300	NDK2 (n=234)		NDK2 (n=206)	NDK2 (n=122) – (only 122 completed both pre and post)	
	WUK4 (n=216)		WUK4 (n=194)		WUK4 (n=9)
	SSP8 (n=210: 130 in community setting & 80 in prison setting) ²⁵			SSP8 (n=210)	
	SSP10 (n=300)			SSP10 (n=300) ²⁶	
	SSP13 (n=170)				SSP13 (n=53 of 157 interviews initiated)
	SSP19 (n=212)			SSP19 (n=212) ²⁷	
	WUK18 (n=213) WUK19 (n=262) CAU5 (n=214)	WUK18 (n=102) WUK19 (n=89)	WUK18 (n=204) WUK19 (n=173) CAU5 (n=214)	WUK18 (n=102) WUK19 (n=?)	CAU5 (n=214)
301-400	WUK15 (n=313)		WUK15 (n=122: 51 in intervention group & 71 in control)	WUK15 (n=122)	WUK15 (T4a n=92: 41 in intervention & 51 in control) and T4b n=62: 27 in intervention & 35 in control)
	CDE1 (n=322) CDE2 (n=424)	CDE1 (n=56)	CDE1 (n=213) CDE2 (n=203)	CDE1 (137) CDE2 (132?)	
401-500	SSP1 (n=451)	SSP1 (n=88)	SSP1 (n=196)	SSP1 (n=108)	SSP1 (T4a n=95, T4b n=89, T4c n=82, T4d n=72, T4e n=50)
	SSP5 (n=410)	SSP5 (n=44)	SSP5 (n=366)	SSP5 (n=366) ²⁸	

²⁵ The first 130 men deemed to have completed community intervention. No attrition reported.

²⁶ Question raised (LO) over reporting of sample, something not quite right. No attrition reported between pre, post & follow-up.

²⁷ Not specified in report but no reported attrition

²⁸ Report does not specify number completed but does suggest any further attrition either?

	SSP6 (n=440 - 364 in intervention group & 76 in control)	SSP6 (n=73)	SSP6 (n=364)	SSP6 (n=266 in intervention group & 44 in control)
501+	SSP14/21 (n=770: 635 in intervention group & 135 control) CAU6 (n=532)	CAU6 (n=312) (non-admitted n=176; drop-outs n=99; excluded n=37)	SSP14/21 (n=598) CAU6 (n=266)	SSP14/21 (n=492: 388 in intervention group & 104 control) CAU6 (n=130 completers only)

Table 5 Information about the sample ²⁹

Information about sample	Pre-start	Start T1	Drop-outs	During Intervention T2		Completion T3	Follow-up T4
				T2a	T2b		
Demographics	SPO1; SSP1/3; SSP5; SSP7; SSP12; SSP14; SSP16; SSP19; WUK11; WEI1; NFI3; CDE3; CCHDE4;	SPO1; SSP1/3; SSP2; WUK2; WUK6; WUK7; WUK15; WUK17; WUK18; CDE1; CDE3; CDE7; CNE1	SSP1/3;	WUK2		CDE1; CCHDE4	WUK15;
Substance misuse, mental health etc ³⁰	SPO1; WUK6; WEI1; CDE3; CNE1	WUK13; SSP2;				SPO1;	
Entry pathway ³¹	SPO1; SSP1/3; SSP4; SSP6; WUK2; WUK4; WUK5; WUK6; WUK7; WUK8; WUK12; WUK13; WUK15; WUK17; WUK18, WEI1; NFI2; CDE3; CCHDE3; CNE1	SSP1/3; WUK4; WUK7				WUK4; WUK6;	
Reasons for dropping out	CNE1;		SSP5; SSP15; WUK1;				

²⁹ NB the info presented here is based on the data extracted i.e. on what is reported - where no info is presented this does not necessarily mean it's not part of the study just that it was not reported. A second round of reading / data extraction will help to standardise the level / type of detail required.

³⁰ Minimum data reported. For many of the Spanish studies drug/alcohol use or mental health problems were reasons for exclusion from the programme so no data available.

³¹ Many of the Spanish studies are court-mandated / alternative to custody so that is main entry path

	WUK7;	
Comparison of prog completers v non-completers?		SSP15;WUK4; WUK7; WUK19; CDE1; CDE2; CDE3; CCHDE1; CAU6; CNE1

3.3 Outcome measures

- Studies across all regions measured outcomes using self-reported data from perpetrators/ participants along with official (CJS) data.
- In all regions except for the South men's self-reported outcome data was also triangulated with data from the victims/partners.
- In all, 30 studies used psychometric instruments (published, standard or validated assessment instruments) to measure change amongst perpetrators/participants, with the large majority of these being used in the Southern studies (n=22). The Southern studies used a battery of validated instruments to collect self-reported psychometric data across a range of measures.
- In some studies (originating from the Western region) psychometric instruments had been developed 'in-house' or existing instruments had been adapted to measure psychological change in perpetrators/participants (e.g. WUK5, WUK18).
- Risk assessments were used in Southern and Western studies as outcome measures (in combination with other measures).
- Change in attitudes towards women, gendered violence and sexist attitudes was measured in studies across North, South, West and Central regions (n=20).
- Western studies covered the widest range of outcome measures, 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.

Table 6 Outcome measures used (all studies by region)

Region	What is being measured?	How is being measured? Instruments used	Who is source of data?
Southern	Attitudes towards women and violence against women; attitudes towards violence; psychological variables; recidivism;	Inventory of distorted thoughts on women (Echeburúa & Fernández-Montalvo, 1998); Inventory of distorted thoughts on violence (Echeburúa & Fernández-Montalvo, 1998); IDT on the Use of Violence; Variables of Mistreatment; SCL-90-R (Derogatis, 1992; González de Rivera, 2002); STAXI-2; STAI; Beck Depression Inventory; Self-esteem scale; Maladjustment scale; Million Clinical Multiaxial Inventory; QVD-VA; Drug & Alcohol Questionnaire; Perceptions of violence scale; Social Desirability Scale 17; Spousal Assault Risk Assessment; Interpersonal Reactivity Index; State-Trait Anger Expression Inventory (STAXI, Miguel-Tobal, Casado, & Cano-Vindel, 2001; Spielberger, 1988); Barratt Impulsiveness Scale (version 10; Barratt, 1985); Rosenberg Self-Esteem Scale; Aggression Questionnaire; Ambivalent Sexism Inventory; CTS-2; AUDIT, CAGE (alcohol use); Dyadic Adjustment scale; Levenson psychopathic scale; Plutchik impulsivity scale; personality disorders, STAXI-2; Social Community Support ASC (Grace, Smith & Musitu, 2002); Global Severity Index; Positive Symptom Total; Positive Symptom Distress Index; BSAFER DV Inventory; Scale of Expectation of Change; Rhode Island Change Assessment; Buss and Dirkee Hostility Index; MIPS; PCL-R; Stressful Life Events Inventory; Support from Close and Intimate Companions Scale; Alcohol Use Disorders Identification Test; Gendered thoughts inventory ((Echeburúa and Montalbo-Fernandez, 2000); (Babor & Grant, 1989). Life Satisfaction. Measured with an item taken from the European Social Survey (2007) Hostile Sexism Scale from the Ambivalent Sexism Inventory (Glick & Fiske, 1996; Expósito, Moya, & Glick, 1998). Community Participation, from the Community Social Support Scale (Gracia & Herrero, 2006b; Herrero & Gracia, 2007a). State Anxiety Scale from the State-Trait Anxiety Inventory (Spielberger, 1988). Anger Control Scale from the State Trait Anger Expression Inventory, STAXI-2 (Spielberger, 1988). Centre for Epidemiologic Studies Depression Scale-7 (Radloff, 1977; CESD short version by Herrero & Gracia, 2007b). Plutchnik Impulsivity Scale (Plutchnik & Van Pragg, 1989). Self-esteem Scale (Gracia, Herrero, & Musitu, 2002). Marlowe-Crowne Social Desirability Scale (Strahan & Gerbasi, 1972). Spousal Assault Risk Assessment-SARA (Kropp, Hart, Webster, & Eaves, 1995; Andrés-Pueyo & López, 2005). Perceived severity of Intimate Partner Violence Scale (Gracia, García, & Lila, 2008). Multidimensional Measure of Emotional Abuse (MMEA, Murphy, Hoover y Taft, 1999), Ambivalent Sexism Inventory, ASI (Glick y Fiske, 1996; versión española de Expósito, Moya y Glick, 1998), Escala de Atribución de Responsabilidad y Minimización (ERM, Lila, Herrero y	Perpetrators/men; Police data; Programme /admin data;

		<p>Gracia, 2008; Lila, Gracia y Herrero, 2012), Quality Marriage Index (QMI, Norton, 1983), Partner Jealousy Index (CR, Montes-Berges, 2008), The State-Trait Anger Expression Inventory (STAXI, Miguel-Tobal, Casado, & Cano-Vindel, 2001; Spielberger, 1988), Personality Questionnaire 17 (Eysenck, Pearson, Easting y Allsopp, 1985; versión española de Luengo, Carrillo de la Peña y Otero, 1991), Cuestionario de personalidad NEO-FFI-R (Costa y McCrae, 1991; versión española de Solé i Fontova, 2006), Aggression Questionnaire, AQ (Buss y Perry, 1992; versión española de Andreu, Peña y Graña, 2002), Escala de Deseabilidad Social (EDS, Crowne y Marlowe, 1960; versión española de Ferrando y Chico, 2000).</p> <p>Police data;</p>	
Western	<p>Repeat offending (perpetrators/men's self-report); pro-domestic violence attitudes; levels of sympathy for battered women (patriarchal attitudes); attitudes towards parenting; motivations and expectations of programme; women's self-reported use of criminal and civil protection orders; women's self-reported quality of life;</p>	<p>Case records; programme monitoring data; observations of & feedback from facilitators; inventory of controlling behaviours (); interviews with men; practitioner's assessment data; police data (crime reports, command & control logs, family protection logs); women's support service records; postal questionnaire; home visits / interviews; own psychometric instrument (adapted from Scott & Wolfe 2000); Sympathy for Battered Women Scale (Saunders 1987) (24 items); Novaco Anger Scale (Novaco 1984) (48 items); Emotional Reliance Upon Others (subscale of Interpersonal Dependency Inventory (Hirschfield 1977) (17 items); Multidimensional Locus of Control (Levenson 1974); Balanced Inventory of Desirable Responding (Paulhus 1984) (impression management scale only); joint session reports; women's reports (interview or written letter); referring professionals reports; info from DV team; CPS data?; women's self-reports; partner abuse scale and parenting abuse scale (not 'validated' developed by Calvin Bell AHISMA Safer Families) both men and women; case studies from case records / programme data; Violence Assessment Index (VAI) (26 items); and Injury Assessment Index (IAI) (21 items); CPS (court records); Risk Interview Schedule for Child Maltreatment RISC; standardised tests Parenting Stress Index; Paulhus Deception Scale; Treatment Motivation Questionnaire; Parental Acceptance Rejection Questionnaire (Rohner & Khaleque 2005), Controlling Behaviour Inventory for Service Users (NSPCC 2007) & Parenting Stress Index Short Form (Abidin 1995); Controlling Behaviour Inventory for Partners (NSPCC 2007) & Adult Wellbeing Scale (Dept Health 2000); Parental Acceptance Rejection Questionnaire (Child) (Rohner & Khaleque 2005), Goodman's Strengths & Difficulties questionnaire (Goodman 1997), Adolescent Wellbeing Scale (Dept Health 2000); official reconviction data from Offenders Index; questionnaire and checklists of abusive behaviours (unspecified); quality of life inventories (unspecified);</p>	<p>Perpetrators/men; Women / partners; Programme facilitators / deliverers; Women's support workers; Police</p>

Eastern	Repeat offending post intervention; health of relationship post intervention; success of intervention (perpetrators/men's self-report – qualitative);	Police recorded charges of gender based violence (GBV); tel interviews with perpetrators/men & partners/family members;	Official data; perpetrators/men; victims (partner / family member)
Northern	Demographic /descriptive data; perpetrators/men; experience of intervention; change in abusive behaviour;	The Conflict Tactics Scale 2 (CTS2), Symptom Checklist (SCL90), KASAM, Audit, Dudit, Client satisfaction questionnaire-8 (CSQ-8) and two self-constructed questionnaires, (1 on the clients experiences of the therapy and 2. gathering descriptive data); Self-constructed questionnaires (on contact with the service, results of the contact regarding violence and results of the contact in general); Clients' change in behaviour, coping with and expressing emotions, responsibility, as perceived by the therapist, and measurements of reductions or changes in physical violence, controlling behaviour, emotional, sexual or other types of abuse (items from CTS); The interaction, design and use of different therapeutic strategies (talking practices of the group participants)measured through discourse and narrative analysis; Self-reported questionnaire capturing recalled behavioural and psychological information from before and after intervention.	Perpetrators/men; women/partners; programme staff/ therapists; official data; programme documentation; 'other' professionals
Central	Socio demographics of perpetrators/men; completion rates; adherence; change in motivation; change in responsibility assumption; relapses during intervention; prognosis; differences between completers & non-completers; what constitutes success; psychological variables; parenting stress; expectations and satisfaction with programme;	Questionnaires; interviews with staff; interviews with professionals; group interviews with perpetrators/men; analysis of perpetrators' journals; group journals; interviews with prosecution authorities; counselling logs; case files;	Perpetrators/men; women/partners; programme data; programme staff; 'other' professionals; official data

Table 7 Measures used

	North (n=8) ³² Sweden, Denmark, Finland, Iceland	South (n=23) ³³ Spain, Portugal	West (n=20) UK, Ireland	East (n=1) Croatia	Central (n=13) Germany, Austria, Switzerland, Netherlands
Measures /instruments used	Interviews/quest with men 4 Interviews/quest with women/partners 1 Interviews /quest with prog facilitators / deliverers 1 Interviews/quest with other / referring professionals 1 'Validated' psychometric instruments ³⁴ 1 Own psychometric instruments ³⁵ 0 Risk assessment 0 Internal / prog data (e.g. admin data, intervention logs, session recordings, case files, perpetrator journals) 2 Official (police / CPS) data 1 Unknown k= 1	Interviews /quest with men 3 Interviews/quest with women/partners 0 Interviews /quest with prog facilitators / deliverers 1 Interviews with other / referring professionals 0 'Validated' psychometric instruments 22 Own psychometric instruments 0 Risk assessment 1 Internal / prog data (e.g. admin data, intervention logs, session recordings, case files, perpetrator journals) 5 Official (police / CPS) data 5	Interviews/quest with men 13 Interviews /quest with women/partners 9 Interviews /quest with prog facilitators / deliverers 6 Interviews with other / referring professionals 12 'Validated' psychometric instruments 5 Own psychometric instruments 2 Risk assessment 1 Internal / prog data (e.g. admin data, intervention logs, session recordings, case files, perpetrator journals) 7 Official (police / CPS) data 8	Interviews/quest with men 1 Interviews/quest with women/partners 1 Interviews /quest with prog facilitators / deliverers 0 Interviews with other / referring professionals 0 'Validated' psychometric instruments 0 Own psychometric instruments 0 Risk assessment 0 Internal / prog data (e.g. admin data, intervention logs, session recordings, case files, perpetrator journals) 0 Official (police / CPS) data 1	Interviews/quest with men 9 Interviews /quest with women/partners 6 Interviews /quest with prog facilitators / deliverers 1 Interviews with other / referring professionals 4 'Validated' psychometric instruments 0 Own psychometric instruments 0 Risk assessment 0 Internal / prog data (e.g. admin data, intervention logs, session recordings, case files, perpetrator journals) 3 Official (police / CPS) data 4 Unknown k= 1

³² Figures changed as of 02.05.14 as 3 Masters theses were excluded from the total number of studies

³³ Figures changed as of 01.04.14 as 6 reports maybe related to only 3 unique studies (SSP1 reports on outcome of study but SSP3 is ex post facto study examining process using the same dataset). Figures in this table assume 23 unique studies in Southern region

³⁴ Published, standard or validated assessment instruments

³⁵ Where the authors have developed their own instruments /adapted others to measure psychological variables

Table 8 What is being measured?

	North (n=8) ³⁶ Sweden, Denmark, Finland, Iceland	South (n=23) ³⁷ Spain, Portugal	West (n=20) UK, Ireland	East (n=1) Croatia	Central (n=13) Germany, Austria, Switzerland, Netherlands
What is being measured ³⁸	Change in attitudes towards women /sexist attitudes 0	Change in attitudes towards women/sexist attitudes 8	Change in attitudes towards women /sexist attitudes 1	Change in attitudes towards women /sexist attitudes 0	Change in attitudes towards women /sexist attitudes 2
	Change in attitudes towards use of violence / gendered violence 1	Change in attitudes towards use of violence / gendered violence 8	Change in attitudes towards use of violence / gendered violence 0	Change in attitudes towards use of violence / gendered violence 0	Change in attitudes towards use of violence / gendered violence 0
	Change in abusive behaviour 0	Change in abusive behaviour 2	Change in abusive behaviour 7	Change in abusive behaviour 0	Change in abusive behaviour 1
	Recidivism /re-conviction (official) 0	Recidivism /re-conviction (official) 5	Recidivism/re-conviction (official) 6	Recidivism /re-conviction (official) 1	Recidivism /re-conviction (official) 3
	Repeat perpetration/ relapse 0	Repeat perpetration/ relapse 0	Repeat perpetration/ relapse 5	Repeat perpetration/ relapse 0	Repeat perpetration/ relapse 6
	Repeat victimisation 0	Repeat victimisation 0	Repeat victimisation 4	Repeat victimisation 0	Repeat victimisation 0
	Quality of life (perp) 0	Quality of life (perp) 2	Quality of life (perp) 1	Quality of life (perp) 0	Quality of life (perp) 5
	Quality of life (of partner/victim) 0	Quality of life (of partner/victim) 0	Quality of life (of partner/victim) 3	Quality of life (partner/victim) 0	Quality of life (partner/victim) 2
	Use of civil & crim protection measures (partner/victim) 0	Use of civil & crim protection measures (partner/victim) 0	Use of civil & crim protection measures (partner/victim) 1	Use of civil & crim protection measures (partner/victim) 0	Use of civil & crim protection measures (partner/victim) 0
	Change in relationship status / quality 0	Change in relationship status / quality 1	Change in relationship status / quality 2	Change in relationship status / quality 1	Change in relationship status / quality 3
	Safety & well-being of victim 1	Safety & well-being of victim 0	Safety & well-being of victim 2	Safety & well-being of victim 0	Safety & well-being of victim 1
	Safety of victim's children 1	Safety of victim's children 0	Safety & well-being of victim 2	Safety of victim's children 0	Safety & well-being of victim 1
	Socio-demographics (perp) 0	Socio-demographics (perp)10	Safety of victim's children 2	Socio-demographics (perp) 0	Safety of victim's children 0

³⁶ Figures changed as of 02.05.14 as 3 Masters theses were excluded from the total number of studies

³⁷ Figures changed as of 01.04.14 as 6 reports maybe related to only 3 unique studies (SSP1 reports on outcome of study but SSP3 is ex post facto study examining process using the same dataset). Figures in this table assume 23 unique studies in Southern region

³⁸ Data extraction for Northern and Central studies in particular not really specific enough to complete some of these counts

Social context of violence 0	Social context of violence 0	Socio-demographics (perp) 2	Social context of violence 0	Socio-demographics (perp) 5
Childhood abuse (perp) 0	Childhood abuse (perp) 0	Social context of violence 0	Childhood abuse (perp) 0	Social context of violence 1
Drug / alcohol use (of perpetrator) 1	Drug / alcohol use (of perpetrator) 4	Childhood abuse (perp) 1	Drug / alcohol use (of perpetrator) 0	Childhood abuse (perp) 0
Completion rates & adherence 0	Completion rates & adherence 2	Drug / alcohol use (of perpetrator) 2	Completion rates & adherence 0	Drug / alcohol use (of perpetrator) 2
Change in motivation 0	Change in motivation 2	Completion rates & adherence 0	Change in motivation 0	Completion rates & adherence 3
Change in responsibility assumption 2	Change in responsibility assumption 2	Change in motivation 1	Change in responsibility assumption 0	Change in motivation 1
Parenting stress 0	Parenting stress 0	Change in responsibility assumption 4	Parenting stress 0	Change in responsibility assumption 1
Other psychological change 3	Other psychological change 13	Parenting stress/skills 3	Other psychological change 0	Parenting stress 1
Differences btw completers & drop-outs 0	Differences btw completers & drop-outs 0	Other psychological change 5	0	Other psychological change 5
Programme expectations / satisfaction (men) 1	Programme expectations / satisfaction (men) 3	Differences btw completers & drop-outs 0	Differences btw completers & drop-outs 0	Differences btw completers & drop-outs 3
Implementation /process 3	Implementation /process 3	Programme expectations / satisfaction (men) 2	Programme expectations / satisfaction (men) 0	Programme expectations / satisfaction (men) 2
Validity of internal eval instruments 0	Validity of internal eval instruments 0	Implementation /process 3	Implementation /process 0	Implementation /process 1
		Validity of internal eval instruments 1	Validity of internal eval instruments 0	Validity of internal eval instruments 1

Table 9 Source of data

	North (n=8) ³⁹ Sweden, Denmark, Finland, Iceland	South (n=23) ⁴⁰ Spain, Portugal	West (n=20) UK, Ireland	East (n=1) Croatia	Central (n=13) Germany, Austria, Switzerland, Netherlands
Source of data	Men's self-report 4	Men's self-report 20	Men's self-report 15	Men's self-report 1	Men's self-report 10
	Official /CJS data 1	Official /CJS data 8	Official /CJS data 8	Official /CJS data 1	Official /CJS data 2
	Women/partners 3	Women/partners 0	Women/partners 10	Women/partners 1	Women/partners 6
	Family members 0	Family members 0	Family members 0	Family members 1	Family members 0
	Children 0	Children 0	Children 1	Children 0	Children 0
	Programme facilitators / deliverers 2	Programme facilitators / deliverers 1	Programme facilitators / deliverers 5	Programme facilitators / deliverers 0	Programme facilitators / deliverers 4
	Programme documentation ⁴¹ 0	Programme documentation 2	Programme documentation 2	Programme documentation 0	Programme documentation 1
	Programme admin 0	Programme admin 2	Programme admin 1	Programme admin 0	Programme admin 0
	Referring professionals / other stakeholders 1	Referring professionals / other stakeholders 0	Referring professionals / other stakeholders 3	Referring professionals / other stakeholders 0	Referring professionals / other stakeholders 0
	Women's service 0	Women's service 0	Women's service 4	Women's service 0	Women's service 0
	Men's self-report only 2	Men's self-report only 12	Men's self-report only 3	Men's self-report only 0	Men's self-report only 2

³⁹ Figures changed as of 02.05.14 as 3 Masters theses were excluded from the total number of studies

⁴⁰ Figures changed as of 01.04.14 as 6 reports maybe related to only 3 unique studies (SSP1 reports on outcome of study but SSP3 is ex post facto study examining process using the same dataset). Figures in this table assume 23 unique studies in Southern region

⁴¹ For example risk assessments or case files

3.4 Follow-up periods

- Over one third of studies (38%) collected data during a follow-up period ($n=25$).
- 5 studies collected data 6 months or less after the programme/ intervention ended (WUK11, SSP6, SSP23, SSP24, CCHDE1)
- 9 studies followed-up between 6 and 12 months after the programme /intervention ended (WUK3, WUK5, WUK6, WUK15, SSP1, SSP10, SSP25, NFI2, CCHDE3)
- 11 studies collected data at the follow-up more than 12 months after the programme /intervention ended (NFI3, NFI4, SSP19, SSP16, SSP13, SSP7/ 12, SSP9, WUK4, CDE7, CAU5, CAU6).

3.5 Impact

57% of the studies ($n=37$) reported moderately positive or promising results (but not statistically significant).

A wide range of impact was reported across the studies:

BEHAVIOURAL CHANGE

- Reduction in *actual* re-offending/ repeated abusive behaviour
- Reduction in the *risk* of re-offending /repeated abusive behaviour
- Change in the type of abusive behaviour perpetrated
- Reduction in the risk of victimisation

ATTITUDINAL CHANGE

- Change in attitudes towards women
- Change in attitudes towards the use of violence (towards women or in general)

PSYCHOLOGICAL CHANGE

- Improvements in (wider) psychological symptoms associated with re-offending or the risk of re-offending

CHANGE IN OTHER RISK FACTORS

- Improvements in other risk factors associated with re-offending/ the risk of re-offending (e.g. mental health, substance misuse, quality of life or life coherence)

MOTIVATION AND ADHERENCE

- Findings regarding motivation / retention / expectations /completion /adherence (e.g. focus of many of the 'central' region studies)
- Findings regarding more general levels of satisfaction / positive experience of intervention (male participants)

Table 10 Reported impact by region

	North (n=8)	South (n=23) ⁴²	West (n=20)	East (n=1)	Central (n=13)
	Sweden, Denmark, Finland, Iceland	Spain, Portugal	UK, Ireland	Croatia	Germany, Austria, Switzerland, Netherlands
Does have an impact?	Yes, statistically significant 0	Yes, statistically significant 11	Yes, statistically significant 1	Yes, statistically significant 0	Yes, statistically significant 1
	Yes, moderate or promising positive (but not statistically significant) 4	Yes, moderate or promising, positive (but not statistically significant) 10	Yes, moderate or promising positive (but not statistically significant) 14	Yes, moderate or promising positive (but not statistically significant) 1	Yes, moderate or promising positive (but not statistically significant) 8
	No 2	No 0	No 0	No 0	No 2
	N/A ⁴³ 1	N/A 0	N/A 4	N/A 0	N/A 1
	<i>Unknown</i> 1	<i>Unknown</i> 2	<i>Unknown</i> 1	<i>Unknown</i> 0	<i>Unknown</i> 1

⁴² Figures changed as of 01.04.14 as 6 reports related to only 3 unique studies (SSP1 reports on outcome of study but SSP3 is ex post facto study examining process using the same dataset). Figures in this table assume 23 unique studies in Southern region

⁴³ Not applicable - implementation /process evaluation

3.6 Limitations

- Lack of use of a control group and small sample sizes in many studies limited the potential conclusions that could be drawn around programme effectiveness.
- In most regions (all except the South) studies sought to explore independent perspectives or observations of third parties during follow-up, which can be of significance for those studies exploring rates of recidivism. However, within the Southern region there existed an over-reliance of perpetrator self-report data, where data from victims/partners could not be accessed. This presented challenges, for example, in understanding the potential effects of social response bias – especially in cases in which participation is mandatory, and parole linked to apparent completion of the programme.
- A further significant limitation consistent across a number of studies was the way in which they dealt with attrition. It seems that some studies executed a ‘treatment-on-the-treated’ analysis that ignored drop-outs and only analysed data on those participants that completed the study. This type of analysis is likely to overestimate the effectiveness of a programme to the extent to which those that dropped out display different characteristics to those that finished the programme, such as levels of motivation etc.

3.7 Standard of reporting

- The quality of reporting (encompassing the level of detail and clarity of articles in explaining study design and results etc) varied greatly across all regions.
- The extent to which programme implementation and contextual information was reported also varied greatly. For example, across the Southern studies, although most papers briefly described contextual issues, including prevalence and incidence of domestic violence and recent government legislation, most papers did not give a detailed account of how the programme fitted within the wider local responses to domestic violence and it was often unclear the precise nature of referral, especially around court-mandated attendance.

4. Summary and conclusion

Our review found an impressive collection of studies from across Europe that perhaps challenges the assumption, including those held by Akoensi *et al* (2013), of a paucity of research in this field. There was substantial use of pre-post design to explore change across time but a lack of use of control groups in the majority of studies limits the potential conclusions that can be drawn about effectiveness. Some of the most methodologically rigorous studies were the most recent ones, which perhaps suggests a move in the right direction.

Southern studies

The approaches to evaluation and methods employed across the Southern studies were very similar; there was a substantial use of pre-post design but still a lack of control groups generally. Studies concentrated on measuring recidivism and psychological change. They employed strict selection criteria and excluded perpetrators presenting mental health and substance misuse problems (i.e. those perpetrators with perhaps the most complex needs). There was one RCT but the sample size was small and thinly spread across the 3 arms / groups. Most samples were drawn from the judicial system and there was high attrition reported. An issue in this region in particular is the lack of validation of the outcomes using data from women/ victims. An initial synthesis of results suggests programmes are effective in addressing certain psychological factors associated with domestic violence, including attitudes towards women, stress, anxiety and depression.

Western studies

Studies originating from the UK and Ireland largely consisted of qualitative studies, with very few using statistical analysis. A number of studies were process-oriented. The outcome studies were very good at triangulating their data sources i.e. using women/victims data and data from other sources to validate the self-reported change amongst participants, in fact no studies relied purely on the self-reports of the perpetrators. On the whole the evaluation results were promising (positive but not statistically significant) particularly in terms of positive changes in the quality of life for the victims/ partners (and their children). Evaluations tended to focus on the wider impact, rather than purely on measuring recidivism – for example changes in quality of life, health of relationships, parenting stress etc

Northern studies

Studies were largely non-experimental, with a focus on participants' experiences of intervention and change in behaviour e.g. coping with and expressing emotions, responsibility assumption (as perceived by the therapist) and measurements of reductions or changes in physical violence, controlling behaviour, emotional, sexual or other types of abuse (items from Conflict Tactics Scale). Studies concentrated on the process of therapy as a vehicle for change, e.g. to make perpetrators take responsibility for their violence. One study relied on men's self-reports only (Sweden). There were some positive outcomes /results but these were not really statistically significant.

Central studies

The evaluation samples consisted of both voluntary and court-mandated participants. Studies from this region were mostly non-experimental, using a combination of quantitative and qualitative data, and robust triangulation of data sources. As with most of the studies in other regions there was a general lack of use of control group design. The outcomes measures tended to focus on issues surrounding programme completion, adherence, change in motivation levels of participants and some analysed the differences between completers and drop-outs. We found one RCT (CNE1) which tested the relative effectiveness of integrated treatment for substance abuse and partner violence to cognitive behavioural treatment addressing substance use disorders among patients in substance abuse who repeatedly committed intimate partner violence (IPV).

Eastern studies

Only one study from this region was found (EHR1). The study used a non-equivalent comparison group design to determine whether, and to what extent, the completion of the Society for Psychological Assistance (SPA) programme for domestic violence perpetrators reduced recidivism. Analysis of police charge data was used to measure recidivism, defined as repeated charges of gender-based violence filed by police in minimum period of 6 months after completion of intervention. Assessment of the health of their relationships post intervention and the 'success' of the intervention was measured via self-reports of men/perpetrators and victims/partners/family members (again conducted at least 6 months after intervention ended). The study results suggested that following programme completion the recidivism rate was significantly lower in the intervention group compared to non-intervention group; there were improved relationships between perpetrators and victims/partners; and 85% of perpetrators and 66% of victims/partners considered the intervention to have been 'successful'. The report however was unclear as to what this 'success' looked like. No demographic data or referral pathway information was reported.

In terms of voluntary versus court-mandated programmes, across the Southern studies many samples were drawn from the judicial system. Research suggests that court-mandated participation, where attendance is backed up by "swift and certain" sanctions for non-attendance or drop-outs, can work because programme completers are less likely to re-offend.⁴⁴ However, such mandated programmes only work with the minority of offenders who have been charged or convicted and fail to reach perpetrators who have not yet come to the attention of the criminal justice system (thus only addressing the tip of the iceberg).

In contrast to the Southern studies, in the Western region participation in most of the programmes evaluated was based on voluntary referrals. Research suggests that perpetrators who self-refer to programmes for reasons such as gaining access to their children or fear of losing their partner (extrinsic motivation) may not be motivated by a genuine desire to change. Conversely, these reasons may also make it more likely that a man will engage on the programme for a longer period. The evidence suggests that motivational factors to attend intervention may change over time, and where a perpetrator is engaged on a programme for a longer term, the more likely they are to experience a change in motivation, developing a more intrinsic desire to change or control their behaviour.⁴⁵ The overview highlighted that more work needs to be done in terms of establishing good quality baseline data, for example, the levels and types of motivation of men/participants.

Health warnings are needed around the levels of attrition reported. Indications of high attrition could be due to the reporting being unclear or inaccurate rather than actual attrition (for example those studies that analyse or report on programme completers only).

⁴⁴Munro T. (2011), *Domestic Abuse Report 5: Conditional Cautioning & Male Perpetrator Programmes*. Report completed as part of 2010-11 Fulbright Police Research Fellowship awarded by the US-UK Fulbright Commission.

⁴⁵. Stanley N. et al. (2011), *Strength to Change: Report of the evaluation of a new initiative for perpetrators of domestic violence*.

Where studies have more of a narrow focus on offenders in terms of applying strict criteria to selection they may not accurately reflect the conditions under which programmes normally operate. This presents a risk of over estimating intervention effect because it ignores or excludes offenders with much more complex issues and needs.

The use of psychometric instruments was widespread across the regions and most prominent within the Southern studies. However, where there is an over-reliance on self-reported data there are likely to be problems of social response bias. One solution to counter this would be to test for social desirability responding.

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.

There exists a varying degree of methodological quality across the literature, with some of the most rigorous being the most recent. Where studies are similar within regions (study designs, outcome measures, etc) the similarities may also be extended to the limitations, suggesting the need for broader dialogue across contexts.

Many of the studies that attempted to measure recidivism encountered significant issues, around the lack of a control /comparison group and the reliability and validity of data. Relying on self-reported rates of recidivism or incomplete police data are likely to underestimate recidivism.

The comparative nature of this exercise highlights an important need for clear and consistent reporting of studies so that the evidence base may be more accessible and applicable to the work addressing domestic violence.