

International Journal of Offender Therapy and Comparative Criminology

<http://ijo.sagepub.com/>

Domestic Violence Perpetrator Programs in Europe, Part II: A Systematic Review of the State of Evidence

Thomas D. Akoensi, Johann A. Koehler, Friedrich Lösel and David K. Humphreys
Int J Offender Ther Comp Criminol 2013 57: 1206 originally published online 3
December 2012

DOI: 10.1177/0306624X12468110

The online version of this article can be found at:
<http://ijo.sagepub.com/content/57/10/1206>

Published by:



<http://www.sagepublications.com>

**Additional services and information for *International Journal of Offender Therapy and Comparative
Criminology* can be found at:**

Email Alerts: <http://ijo.sagepub.com/cgi/alerts>

Subscriptions: <http://ijo.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://ijo.sagepub.com/content/57/10/1206.refs.html>

>> [Version of Record](#) - Sep 16, 2013

[OnlineFirst Version of Record](#) - Dec 3, 2012

[What is This?](#)

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
© The Author(s) 2012
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0306624X12468110
ijo.sagepub.com



Thomas D. Akoensi¹, Johann A. Koehler¹,
Friedrich Lösel^{1,2}, and David K. Humphreys¹

Abstract

In Part II of this article, we present the results of a systematic review of European evidence on the effectiveness of domestic violence perpetrator programs. After searching through 10,446 titles, we discovered only 12 studies that evaluated the effectiveness of a perpetrator program in some systematic manner. The studies applied treatment to a total of 1,586 domestic violence perpetrators, and the sample sizes ranged from 9 to 322. Although the evaluations showed various positive effects after treatment, methodological problems relating to the evaluation designs do not allow attribution of these findings to the programs. Overall, the methodological quality of the evaluations is insufficient to derive firm conclusions and estimate an effect size. Accordingly, one cannot claim that one programmatic approach is superior to another. Evaluation of domestic violence perpetrator treatment in Europe must be improved and programs should become more tailored to the characteristics of the participants.

Keywords

domestic violence, systematic review, intervention, cross-cultural, rehabilitation

¹University of Cambridge, UK

²University of Erlangen–Nuremberg, Erlangen, Germany

Corresponding Author:

Thomas Akoensi, Institute of Criminology, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, UK.

Email: tda24@cam.ac.uk

Introduction

In Part I (Hamilton, Koehler, & Lösel, 2013) of this article, we provided the results of an international survey of domestic violence perpetrator programs in Europe. The results of that review indicated that measures to reduce the incidence of episodes of future reabuse were common throughout the continent and that most European countries had at least some measure in place to deal with the abusive attitudes and behaviors of violent men. However, we also observed a wide range of program designs and implementation practices, and a particular dearth of high-quality evaluation integrated into routine practice. It is thus necessary to establish how closely practitioners in Europe base their routine practices on a definitive body of up-to-date knowledge concerning the effectiveness of domestic violence perpetrator programs.

However, for various reasons, the empirical evidence guiding practitioners in Europe is unsatisfactory. First, scholars generally note a lack of conclusive evidence concerning perpetrator program effectiveness, which seems to relate to various methodological factors pertaining to evaluation standards (Davis & Taylor, 1999; MacKenzie, 2006; Rosenbaum, 1988). For instance, North American evaluations of the effectiveness of various perpetrator programs remain thus far largely inconclusive (e.g., Babcock, Green, & Robie, 2004; Davis & Taylor, 1999; Feder, Wilson, & Austin, 2008; Hamberger & Hastings, 1993; MacKenzie, 2006). In a meta-analytic examination of the effectiveness of court-mandated offender programs, for example, Feder et al. (2008) observed reductions in reoffending of 13% among participants who had undergone treatment. However, these improvements disappeared when the researchers examined victim-reported outcomes. Similarly, Babcock et al. (2004) observed small overall treatment effects from perpetrator programs that diminished further when they applied stringent methodological criteria in their analyses. Moreover, the researchers found no significant difference in effectiveness when comparing different treatment modalities.

Second, the ongoing debate among scholars concerning the correct interpretation of the etiology and treatment of domestic violence perpetration (e.g., Dutton & Corvo, 2007; Gondolf, 2007) leaves practitioners with little guidance of how best to deliver rehabilitative service. As yet, the dearth of evidence emphasizing the superiority of one treatment modality over another means that the validity of these theoretical contributions continues to remain unresolved. Although we observed in our survey that practitioners express a preference for a variety of program designs and delivery styles, it is not clear whether treatment components from one approach militate against the effectiveness of treatment components from another. Consequently, the lack of evidence-based guidance concerning “what works” has especially salient ramifications concerning the type and quality of treatment that is delivered to offenders.

Third, caution should be exercised when generalizing the results of the North American reviews that were mentioned earlier to a European context, given the highly politically charged and culturally embedded nature of domestic violence (Gracia & Herrero, 2006). Populations of abusers, legal frameworks, and treatment approaches may have unique implications for the delivery of perpetrator treatment.

Despite the presence of perpetrator treatment programs on the domestic violence landscape for more than 30 years (Barnish, 2004), the significant developments in research on these interventions remain thus far in large part theoretical. Therefore, given that there has been no synthesized body of European evidence on “what works” for domestic violence perpetrator programs, practitioners have to draw on often unconnected research from countries with differing issues of transferability. It is in this context that we present below the results of a systematic review of European evidence on the effectiveness of domestic violence perpetrator programs.

Method

Our literature search contained the following eligibility criteria:

1. *Region*: We limited our search to evaluations conducted in Europe. This was not restricted to European Union (EU) countries alone.
2. *Target population*: The target sample had to comprise domestic violence perpetrators, defined as either offenders who had been convicted of a domestic violence offense, or partners who had commenced a course of treatment to deal with their self-reported partner-abusive behavior. Programs targeting male and female perpetrators were eligible.
3. *Intervention*: The evaluation had to examine the effectiveness of a treatment program that was designed to alter the attitudes and/or behaviors of domestically violent partners.
4. *Evaluation design*: At minimum, the study had to measure outcomes before the commencement of treatment, and at the conclusion of treatment, corresponding to Level 2 on the Maryland Scale of Scientific Methods (Sherman, Farrington, Welsh, & Mackenzie, 2002). Due to a lack of methodologically more sound evaluations (e.g., randomized controlled trials, matched-pairs designs), we had to apply this fairly relaxed criterion.
5. *Outcomes*: Both attitudinal and behavioral outcome measures were acceptable.
6. *Publication*: Both published and unpublished formats were acceptable for inclusion.
7. *Language*: Studies could be written in any common European language.

To locate unpublished and published studies, we searched online computerized databases and specialist journal archives,¹ as well as meta-analytic and systematic review publications dealing with domestic violence perpetrator programs (e.g., Babcock et al., 2004; Davis & Taylor, 1999; Feder et al., 2008; Rothman, Butchart, & Cerdá, 2003). We also contacted academics and experts in an effort to locate studies that might not have been accessed by the more conventional strategies. In addition, as part of a corresponding survey of domestic violence perpetrator programs throughout the EU (see Part I of this article), we asked respondents to furnish us with any evaluations of their

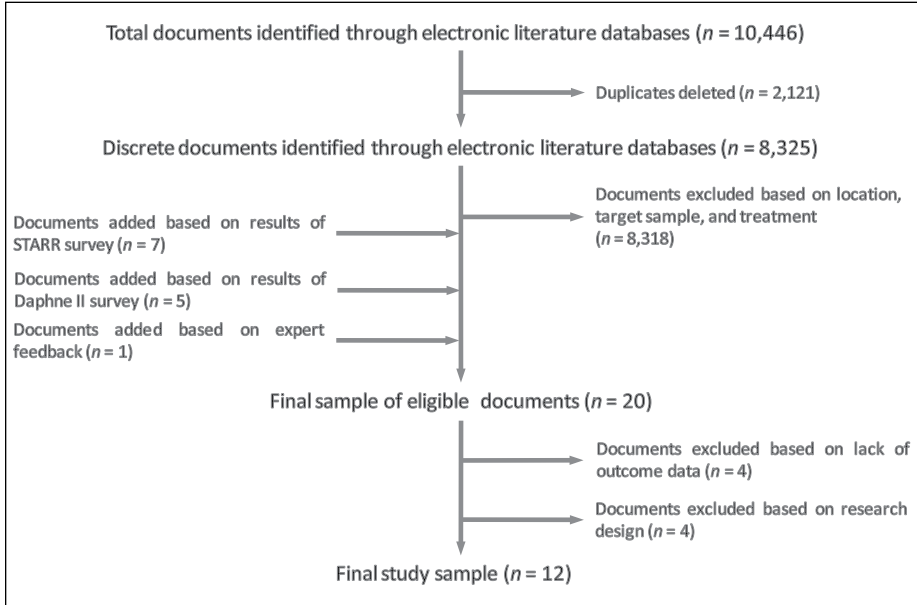


Figure 1. Flowchart of literature search.

practice. Moreover, we consulted the database of the Daphne II Work With Perpetrators Survey, which had compiled a network of domestic violence perpetrator programs in each of the 27 EU countries during 2007 to 2008. Although this database did not include information on outcome evaluations, we individually contacted each respondent to that survey and asked them to provide us with any available evaluations of their program. Figure 1 provides an illustration of the search process.

Our bibliographic database search yielded a total of 10,446 titles, which, on deletion of duplicates, yielded 8,325 discrete documents. The titles and abstracts were then screened according to method, location of study, sample population, and outcome of interest, to arrive at 7 studies. These were supplemented by 8 further studies that were retrieved in our questionnaire survey and from expert feedback, and a further 5 studies were added through consultation with respondents to the Daphne II Work With Perpetrators Survey. This resulted in 20 studies, which we retrieved in full. Eight studies were excluded on the basis of a lack of outcome measurements ($k = 4$), or because measurements were taken at only one point in time ($k = 4$). Our final study sample consisted of 12 evaluations that fulfilled all eligibility criteria.

Results

The 12 studies originated from six European countries: Cyprus ($k = 1$), Finland ($k = 1$), Germany ($k = 1$), Spain ($k = 4$), Sweden ($k = 1$), and the United Kingdom ($k = 4$).

Six studies were published, and 6 were unpublished.² Ten studies were published between 2000 and 2010; the oldest was written in 1997.

We provide an in-depth narrative review below of the seven studies that report outcomes related to offending or violent behaviors. Five studies in our sample used data that pertained to attitudes and beliefs surrounding women and psychological variables related to impulsivity, self-esteem, anger, and so on. We eschew in-depth description of these studies and instead refer readers seeking elaboration to Table 1 for further details.

R. P. Dobash, Dobash, Cavanagh, and Lewis (1999)

R. P. Dobash et al. (1999) compared the effectiveness of two court-mandated perpetrator programs with traditional criminal justice-based sanctions (e.g., fines, probation, and prison) in Scotland. They were cognitive-behavioral and emphasized educational rather than psychodynamic methods. The authors administered interviews to the men and the women partners of the participants at the beginning of the program (Time 1) and sent postal questionnaires at 3 and 12 months thereafter (Times 2 and 3). There were few differences between the two groups on key demographic, criminal, and attitudinal variables at Time 1, although there were significant differences pertaining to employment and marital status. During the follow-up period, marginally more men in the treatment condition appeared in arrest and prosecution records than men in the control condition. Women's reports of subsequent violence based on questionnaire data revealed that 33% of the men in the experimental condition and 69% in the control condition used violence at Time 3. This difference was statistically significant. A similar difference obtained when the authors compared the use of frequent violence between the two groups. The authors also observed reductions in the experimental condition of controlling and intimidating behaviors, both over time and compared with the control condition. Moreover, women partners of men in the experimental condition reported more positive and statistically significant improvements in quality-of-life measures such as feelings of happiness, contentment, and safety than women partners of men in the control condition.

Bowen (2004)

Bowen (2004) investigated the effectiveness of a court-mandated program delivered by the West Midlands Probation Area in the United Kingdom. Attendance at fewer than 21 of the 24 core sessions constituted a dropout. The modules adopted a psycho-educational and profeminist approach. Outcomes were measured by official police records as well as by means of a number of psychological batteries administered at 11-months follow-up, which captured data about violent behaviors and attitudes surrounding anger, violence, and dependency. Psychological variables were adjusted to compensate for social desirability bias. In the 11-months follow-up period, completers were alleged to have committed fewer domestic violence offenses than dropouts.

Table 1. Summary of Primary Studies.

Author(s)	Country	Participation	Theoretical paradigm	Program structure	Sample size and drop-out rate	Maryland scale	Outcome measure and follow-up period	Results
R. P. Dobash, Dobash, Cavanagh, and Lewis (1999)	United Kingdom	Court mandated	Duluth model; cognitive-behavioral therapy (CBT)	Two group programs: 24 weeks and 27 weeks	51 men and 47 women in treatment group, 71 men and 87 women in comparison group. In all, 47% male and 40% female dropout in treatment group, 51% male and 42% female dropout in comparison group.	3	Various psychometric and psychological assessment instruments, official crime reports, and women partner's self-reports. Outcomes measured at intake and 3 and 12 months after program completion.	7% recidivism in treatment group, 10% in comparison group; reductions in violence corroborated by women partner's reports. Improvements in treatment group's men's controlling behaviors and women's well-being, compared with control group.
Bowen (2004) ^a	United Kingdom	Court mandated	Profeminist psycho-educational (Duluth model)	Twenty-four 150-min group sessions, and five 150-min follow-up sessions.	120 men, 32% dropped out.	2	Various psychometric and psychological assessment instruments and official crime reports. Outcomes measured at intake and at 11-months follow-up.	15% recidivism for treatment group and 33% for dropouts, equating to a small, marginally significant effect. Offenders achieved modest psychological improvements, although these were not related to offending behaviors.
Leicester-Liverpool Evaluation Group (2005) ^a	United Kingdom	Court-mandated Community Rehabilitation and Punishment Order (CRPO)	Duluth model	Twenty-four 2-hr sessions over 6 months	262 men, of whom 89 did not begin treatment. Of those who commenced treatment, 45.1% dropped out.	2	Official reconviction measure based on the Offender Index, collected 8 months after completing treatment. Behavioral and psychometric data were collected from program completers and their partners before and after treatment.	50.4% of the sample were reconvicted, among whom 28 (29.2%) had completed treatment, 55 (70.5%) dropped out, and 49 (55.1%) did not begin the program. Program completers and their partners reported positive change on behavioral and psychometric measures.

(continued)

Table 1. (continued)

Author(s)	Country	Participation	Theoretical paradigm	Program structure	Sample size and drop-out rate	Maryland scale	Outcome measure and follow-up period	Results
Adva (2008) ^a	United Kingdom	Voluntary	Duluth model; CBT	10 individual CBT sessions; 30 loosely defined group sessions.	115 men, 63% dropped out. Data collected from 12 women partners and 20 children of perpetrators.	2	Self-reported risk of reabuse, psychological variables (e.g. self-esteem, locus of control), abusive incidents, measured throughout treatment over rolling period of 30 months.	Initial increase in self-reported abusive behavior; then gradual decline. Women partner report decrease in abuse. Strong decrease in risk of reabuse among treatment completers, corroborated by women's report. Significant psychological improvement among perpetrators and among women and children.
Societystyrelsen (2010)	Sweden	Voluntary	Duluth model; CBT	Varied from 3 individual to more than 20 individual/group sessions.	188 men, 43% dropped out. 16 women partners, 25% dropped out.	2	Self-assessment instrument capturing various psychologically and physically violent behaviors, mental health, and substance abuse, captured at Time 1 (entry into program) and at Time 2 (12 months thereafter). Men's and partner's views of treatment gathered at Time 2.	Continuation of various forms of violent behaviors, although some reductions observed in the sample. Partners reported reductions in violent behaviors. Improvement in perpetrators' mental health and substance abuse. Perpetrators and partners reported satisfaction with the program.
Törmä and Tuokkola (2009) ^a	Finland	Voluntary and court mandated	Psychodynamic	Individual or group sessions. Treatment lasted at minimum 2 months.	80 men responded to questionnaire. Six women spouses of perpetrators.	2	Self-report questionnaire capturing recalled behavioral and psychological information from before and after treatment. Interviews with women partners.	Reported reductions of violent incidents, intimidating behaviors, and an increase in well-being. Spouses reported feeling safer as a result of men's participation in treatment program.

(continued)

Table 1. (continued)

Author(s)	Country	Participation	Theoretical paradigm	Program structure	Sample size and drop-out rate	Maryland scale	Outcome measure and follow-up period	Results
Association for the Prevention and Handling of Violence in the Family (2009) ^a	Cyprus	Not specified	Psycho-educational	Twenty 2-hr seminars over 12 weeks.	9 men	2	Questionnaire measuring psychological variables at intake and on program completion.	Modest psychological improvements in anger, well-being, anxiety, and self-esteem, and modest reductions in physical assault toward partners.
Hagemann-White, Kavemann, and Beckmann (2004)	Germany	Voluntary and court mandated	CBT	Roughly 20 group sessions	322 men, 57% dropped out.	2	Self-report questionnaire capturing recalled behavioral and psychological information from before and after treatment.	Improvements in gender-stereotypical attitudes, accepting responsibility, and admitting guilt.
Echeburúa & Fernández-Montaño (2009)	Spain	Voluntary	CBT	20 weekly 2-hr group sessions.	148 men, 32% dropped out.	2	A battery of psychological assessment instruments, administered at intake and immediately after completion of the program.	Promising change on various psychological variables related to impulsivity, anger, and self-esteem.
Echeburúa, Fernández-Montaño, and Amor (2006)	Spain	Voluntary	CBT	20 weekly 2-hr group sessions.	52 men, 8% dropped out.	2	A battery of psychological assessment instruments, administered at intake and immediately after completion of the program.	Promising change in cognitive distortions, hostile attitudes, and uncontrolled anger.
Echeburúa and Fernández-Montaño (1997)	Spain	Voluntary	CBT	15 weekly 1-hr individual sessions	31 men, 48% dropped out.	2	A battery of psychological and psychometric assessment instruments, administered at intake, end of treatment, and at 1- and 3-months follow-up.	Promising change in various psychological variables, for example, anxiety, self-esteem, depression, anger.
Echauri Tijeras (2010) ^a	Spain	Voluntary and remand	CBT	20 weekly 2-hr group sessions.	170 outpatient men, 53% dropped out; 80 men in prison, 73% dropped out.	2	A battery of psychological assessment instruments, administered at intake, end of treatment, and at 1-, 3-, 6-, and 12-months follow-up.	Promising change in various psychological variables, for example, anxiety, self-esteem, depression, anger.

^aThe study was unpublished.

Recidivism was not associated with risk, offender type, the therapeutic environment, program attendance, or variations in program implementation. Reoffending was significantly associated with pretreatment criminal history. Although a small positive psychological change was observed in the offender sample, this was not related to reoffending.

Leicester–Liverpool Evaluation Group (2005)

The authors evaluated the effectiveness of a Duluth-based perpetrator program applied to 173 men in three sites across the United Kingdom. Treatment comprised twenty-four 2-hr sessions, delivered in 6 months, to perpetrators who had been subjected to a Community Rehabilitation and Punishment Order. Assessments of the perpetrators' reconviction patterns were collected from the Offenders Index, and behavioral and psychometric data were collected from program starters and completers.³ The authors observed lower reconviction rates among treatment completers than among noncompleters and perpetrators who had not commenced treatment. They also observed significant improvements in perpetrators' and partners' self-reported abusive behaviors, on completion of the program.

Adva (2008)

The authors investigated the effectiveness of a community-based treatment in the United Kingdom that targeted perpetrators who had been referred by other domestic violence agencies or who had approached the service directly. The intervention comprised meetings with the police domestic abuse unit, social workers, and women's and children's workers. The treatment consisted of a 42-week course, involving 10 individual cognitive-behavioral therapy sessions, and 30 loosely defined group sessions with the perpetrator. The modules adopted cognitive-behavioral and Duluth-based psychodynamic methods. Continuous assessment was undertaken throughout the duration of participation in the treatment. The percentage of men self-reporting abusive behavior increased sharply after the 1st month, and then decreased over the next 11 months until it was somewhat lower than at the commencement of assessment. The risk level of abuse decreased for the majority of participants who completed treatment, and this change was less pronounced among participants who did not complete the program. Furthermore, this was corroborated by the women partners' perceptions of the risk of abuse. Statistically significant improvements on a number of psychological variables were observed among 19 treatment completers. Partners of program participants reported a decline in the number of abusive incidents over the course of the intervention and an improvement in well-being and safety regardless of the man's progress through treatment. The majority of the 20 children of perpetrators who were assessed toward the beginning and end of their support program reported psychological, behavioral, and academic improvements.

Socialstyrelsen (2010)

This study investigated the effectiveness of eight voluntary programs located throughout Sweden. The programs were broadly similar, and comprised individual and group sessions incorporating psychodynamic and cognitive-behavioral approaches. The length of treatment varied for each participant from 3 individual sessions to more than 20 individual and group sessions each week. A questionnaire was administered at entry into the program (Time 1) and at 12 months thereafter (Time 2) to 188 male participants and 16 female partners. Questionnaire items concerned physical and psychological violence, mental health, and substance abuse. At Time 2, data were gathered from 12 women and from 140 program participants. The author used a “last observation carried forward” analysis for the male participants who were not available at posttest measurement. At Time 2, 38% of participants had completed treatment, 43% had dropped out, and 19% were still in treatment. The assessment comprised dichotomous-level data capturing whether or not a violent behavior had been used in the year preceding measurement. Although there were statistically significant reductions in the use of various forms of violence at Time 2, the majority of men continued to use minor psychological violence, and a substantial proportion reported continued physically violent behaviors. Significant improvements were observed in program participants’ mental health and substance abusive behaviors at Time 2, compared with Time 1. The majority of participants reported satisfaction with the program at Time 2. Eleven (92%) of the women reached at Time 2 who were still in contact with their partners reported that violent behaviors had decreased.

Törmä and Tuokkola (2009)

Törmä & Tuokkola investigated the effectiveness of a treatment program in Finland that accepted either court-mandated or voluntary referrals, and that involved one-on-one or group discussion therapy depending on client intake assessments. The intended duration of the psychodynamic treatment was at minimum 2 months. The authors distributed outcome evaluation surveys to selected clients who had participated in the program; however, this selection process was not described in detail. The time that elapsed between program completion and questionnaire distribution varied among respondents. In all, 61% ($n = 80$) of questionnaire recipients responded, providing data about violent behaviors, psychological change, and satisfaction with the program, based on recollections from before and after participation. The authors reported reductions in self-reported “violent incidents,” “intimidating behaviors,” and increased “well-being” across all respondents. Many respondents reported that the program had “been useful” to them. The authors also interviewed six spouses of program participants, all of whom reported “feeling safer” as a result of the men’s participation in the program. The precise behaviors, attitudes, and beliefs captured within these constructs were not specified.

Association for the Prevention and Handling of Violence in the Family (APHVF; 2009)

The authors investigated the effectiveness of a community-based perpetrator program delivered to domestically violent men in Cyprus. The program comprised 22-hr psycho-educational group sessions delivered over the course of 12 weeks. Nine participants completed a questionnaire at intake into the program and immediately on program completion. The questionnaire captured the respondents' sense of anxiety, well-being, anger, and self-esteem, as well as instances of psychological and physical aggression toward the partner. Modest improvements were observed among all psychological variables, and five of the six participants who had reported physical assault at intake reported a reduction upon program completion. Of the three cases in which respondents reported injuring their partner at intake, reductions were reported in two cases. The authors conducted no statistical analysis of the data.

Discussion

We intended this review to provide the first synthesis of European evaluations of the effectiveness of domestic violence perpetrator programs. Given our prior lack of knowledge about the state of evaluation research in this field in Europe, we relaxed the methodological eligibility criteria to retrieve a study sample that accommodated a variety of program and research designs. Our approach uncovered a substantial body of recent evaluations, emanating from diverse sources. A considerable number of primary studies in the sample were from countries that are traditionally underrepresented in the criminological literature or were retrieved from sources that would have been concealed had we relied on conventional search strategies (Wilson, 2009). When applying a crude "vote count" method of surveying the outcomes across the 12 studies in our sample, various positive effects were observed in all cases (see Table 1).

However, the quality of the studies within our sample do not, in aggregation, support confident claims about how well such programs work, and under what circumstances practitioners can expect to reduce the incidence of future episodes of reabuse. In the following, a few of the problems will be briefly discussed.

Evaluation Design

There was a near-total absence of comparison group evaluation designs in our sample. Of the 12 studies, only 1 used such a design (R. P. Dobash et al., 1999). However, the equivalence between the two groups in that study was questionable because the participants were allocated to groups based on their court sentence. The authors note that participants may have been mandated to attend the programs only because the local law enforcement officials deemed them to be less severe offenders (R. P. Dobash et al., 1999). It is therefore plausible that this exaggerated the effectiveness of the treatment.

The remaining studies gathered outcome data from only those participants who had undergone treatment. Although all 12 studies reported reductions in one or the other outcome measures, one cannot draw a firm causal conclusion. For example, various threats to internal validity (e.g., history, maturation, selection, and statistical regression) apply to within-group evaluation findings (Lösel, 2007; Shadish, Cook, & Campbell, 2002). In particular, the modest to significant reductions may simply be an artifact of the “honeymoon period” phenomenon (Rosenbaum, 1988; see below) and not a true indicator of perpetrator program effectiveness.

Program Type

All 12 studies adopted an approach that mixed cognitive-behavioral, educational, and profeminist techniques. This renders identification of the precise components of effective treatments rather difficult. The specific method of treatment delivery in most cases was not described in detail and thus manifested low descriptive validity (Lösel, 2007; see also Mears, 2003, for perpetrator program research specifically). Moreover, with one exception (Echeburúa & Fernández-Montalvo, 1997), all the treatments comprised group therapy sessions with similar levels of intensity and dosage. Sessions were typically spread over 20 weeks, ranging from 3 to roughly 30 sessions (R. P. Dobash et al., 1999; Socialstyrelsen, 2010, respectively). Consequently, although complex and theoretically heterogeneous program packages seem common, it is difficult to evaluate which treatment components may have led to more or less positive results (if other threats to validity would not have been present). This seems not to be a specific pattern of European research because North American reviews were similarly unable to discern which treatment approaches worked better than others (e.g., Babcock et al., 2004; Feder et al., 2008).

Representativeness of the Samples

The sample sizes in the primary studies of our European review ranged considerably, from 9 to 322 (APHVF, 2009; Hagemann-White, Kavemann, & Beckmann, 2004, respectively). The participants were selected from a range of sources, such as voluntary referrals and court-mandated diversion orders. This heterogeneity leads to the question of the generalizability of the samples of treatment participants. Feder et al. (2008) observed in their North American meta-analysis that programs applied to a general population that was representative of “typical” perpetrators observed lower overall mean effects than programs that were applied to a uniquely appropriate subset of abusers. They attributed this finding to the possibility that such programs may be especially dependent on various factors relating to the participant group, such as levels of motivation to complete treatment (Feder et al., 2008). It is thus necessary to distinguish when the evaluations have isolated those participants who are most likely to manifest reductions in intended outcomes, whether through sample size, selection, or attrition.

Perpetrator Characteristics

Of the 12 studies in our sample, none explicitly tailored the treatment program to specific characteristics of the participants. For example, although Bowen (2004) and R. P. Dobash et al. (1999) conducted analyses on the effectiveness of the treatment on different subtypes of perpetrator, these were post hoc and were not used to match treatment to the individual perpetrators' needs. Perhaps the high drop-out rates resulted from programs having targeted a particular type of perpetrator but included other participants whose criminogenic needs may have been partially different (Day, Chung, O'Leary, & Carson, 2009; Graham-Kevan, 2007). As treatment effectiveness can be increased when program delivery is tailored to participants' learning styles and behavioral profiles (e.g., Andrews & Bonta, 2010; Koehler, Lösel, Akoensi, & Humphreys, 2013; Lösel, 2012), further attention should be paid in the future to an adequate combination of standardization and some individualization of perpetrator programs (see also Cavanaugh & Gelles, 2005).

Attrition

In most studies we observed relatively high rates of attrition. In the prison subsample of Echauri Tijeras (2010), for example, only a quarter of the original sample completed the program. In the other studies, attrition rates were rarely below 30%. This is especially concerning in light of the observation that predictors of dropout seem to be related to perpetrators' risk factors for the resumption of violent behaviors (Daly & Pelowski, 2000). Moreover, one must assume that the women victims lost at follow-up, such as those in R. P. Dobash et al. (1999) and in Socialstyrelsen (2010), are more likely to be abused with greater frequency and severity (Sullivan, Rumptz, Campbell, Eby, & Davidson, 1996). It is highly plausible that the program completers had the strongest motivation to change. A selection bias such as this would have exaggerated the results in favor of finding a reduction in abuse at follow-up. This phenomenon has been labeled *creaming*, and has been observed in previous North American systematic reviews (Babcock et al., 2004; Davis & Taylor, 1999; Feder et al., 2008). Furthermore, the conflation of outcome data for both voluntary and court-mandated samples together (e.g., Echauri Tijeras, 2010; Hagemann-White et al., 2004; Törmä & Tuokkola, 2009) dilutes the effectiveness of the program, as voluntary referrals are also likely to manifest greater motivation to change.

Outcome Measures

A key methodological issue is the appropriateness of outcome measurement for perpetrator program effectiveness (e.g., R. E. Dobash & Dobash, 2000; Stover, 2005; Westmarland & Kelly, 2012). For example, there is continuing disagreement about whether official police data, perpetrator self-reports, victim interviews, or program deliverer testimonies

provide the most reliable indicator of repeat violence, as each form entails unique flaws (Straus, 1991). For this reason, it is generally advised that evaluations incorporate data from a variety of sources, in an attempt to “triangulate” often-inconsistent information (Gondolf, 2002; Rosenbaum, 1988). However, this was done in very few studies in our sample. One study (R. P. Dobash et al., 1999) gathered outcome data from official police report, offender questionnaire, and women partners’ self-reports. Two studies (Bowen, 2004; Leicester–Liverpool Evaluation Group, 2005) corroborated offender self-report data with police records, and two studies (Socialstyrelsen, 2010; Törmä & Tuokkola, 2009) complemented offender self-report data with information gathered from questionnaires administered to the women partners. The remaining seven studies collected outcome data from the offenders’ self-reports alone.

With regard to the *content* of the outcome measure, data can vary from criminal justice measures such as arrest, over rates of physical assault or episodes of verbal abuse, to perceptions of chronic intimidation and a general sense of lack of safety and well-being. Only 3 of the studies in our sample collected official crime data from police records (Bowen, 2004; R. P. Dobash et al., 1999; Leicester–Liverpool Evaluation Group, 2005), and 4 studies collected data concerning self-reported violent behaviors such as slapping and beating. All 12 studies gathered data concerning psychological change in some form or another. However, the utility of such data in determining future violence remains unclear. For example, Bowen’s (2004) analysis of the relationship between offenders’ responses to items capturing psychological change revealed that there is little concordance with reoffending behaviors.

Length of Follow-Up

The available literature has alluded to a “honeymoon period” of either a cessation or subsidence of abusive behaviors in the duration and immediate aftermath of enrollment in a perpetrator program (e.g., Rosenbaum, 1988; Rosenfeld, 1992). As a consequence, the collection of outcome data immediately on program completion risks inflating the probability of asserting the false-positive of an encouraging treatment effect. Half ($k = 6$) of the studies collected data only immediately on program completion. Of the six remaining studies, none collected data more than 12 months after the program had concluded. Some researchers have advocated dismissing results that have been gathered any sooner than 6 months after the treatment program has ended (e.g., Feder et al., 2008). Had we done so, we would have excluded 6 of the 12 studies in our sample.

Other Issues

Three additional features of the evaluations may have influenced the low methodological quality observed in our sample. First, developers or administrators of the programs were prominently involved in many of the evaluations ($k = 5$). This mirrors

routine practice in perpetrator program delivery throughout Europe more generally (see Part I of this article). It is plausible that in these instances, the evaluations were conducted not only as an objective test of the program's effectiveness but rather as an administrative tally of the program's general performance. Second, many of the evaluations ($k = 6$) were conducted in the community, as opposed to custody, where the formation of a control group is encumbered by the difficulty in acquiring perpetrators who are interested in participating in an evaluation. Finally, one must assume cultural influences that generally accompany less openness to sound evaluations of offender treatment in Europe than in North America (e.g., Lösel, Koehler, Hamilton, Humphreys, & Akoensi, 2011).

Conclusion

This project represents the most comprehensive attempt yet to provide an overview of work with domestic violence perpetrators throughout Europe. In Part I of this article, we outlined the scope and variety of routine practice among European perpetrator programs, and in Part II we examined whether the evidence base supported the effectiveness of those approaches. However, because of the above-mentioned and other problems, this systematic review could not reveal definitive conclusions regarding the effective delivery of domestic violence perpetrator programs in Europe. Our findings resonate with the conclusions of reviews from North America, namely, that we do not yet know what works best, for whom, and under what circumstances (e.g., Babcock et al., 2004; Davis & Taylor, 1999; Feder et al., 2008; Hamberger & Hastings, 1993). Compared with other fields of offender treatment (e.g., Lipsey & Cullen, 2007; Lösel, 2012), evaluation research on domestic violence perpetrator programs is still at an early stage.

Appendix A

Search Terms

Subject	Program	Outcome and methodology
Domestic violence	Program(s)	Effect*
Domestic assault	Treat*	Outcome*
Batterer	Intervention(s)	Eval*
Family violence	Therapy	Experiment*
Physical abuse	Counsel*	Randomized controlled trials (RCT)
Spousal abuse	Rehab*	Quasi (experiment*)
Interfamily violence	Court decisions	Trial
Intimate partner violence	Mandated court decisions	Empirical
Duluth	Within prison	Recidiv*

*Search term wildcard.

Appendix B

Databases Searched

Electronic databases	Hand search of gender violence journals	Government publications
International Bibliography of the Social Sciences (IBSS)	<i>Journal of Interpersonal Violence</i>	UK Home Office Research database
PsycInfo	<i>Advances in Psychiatric Treatment</i>	Brå-Swedish National Council for Crime Prevention
PsycArticles	<i>Feminist Criminology</i>	
PubMed	<i>Feminist Theory</i>	
Cochrane Library	<i>Violence against Women</i>	
C2-SPECTR	<i>Journal of Family Violence</i>	
EmBase	<i>The Family Journal</i>	
ISI Web of Knowledge	<i>Feminism and Psychology</i>	
CSA Illumina		

Acknowledgments

We thank our partners from the UK Ministry of Justice, London Probation Trust, European Organization for Probation, and the Ministries of Justice of Bulgaria, France, and Hungary for their cooperation. We graciously acknowledge the assistance provided by the many experts and practitioners who responded to our requests for information and data, and by the Daphne II Work With Perpetrators Survey who furnished us with data that assisted in our literature search.

Authors' Notes

This study was carried out within the European cooperation project "Strengthening Transnational Approaches to Reducing Reoffending."

Declaration of Conflicting Interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The project was funded by the European Commission.

Notes

1. A full list of databases used in our search, as well as the search terms employed, can be found in Appendices A and B.
2. One study (Bowen 2004) was an unpublished doctoral dissertation. Although this work has appeared in published form in peer-reviewed journals, we referred to the original source

document on the basis of its comprehensive explanation of the primary study project and its outcomes.

3. Behavioral and psychometric data were not reported in Leicester–Liverpool Evaluation Group’s (2005) study but were reported in Hatcher et al. (2003). We refer to the former document, as it provides the most recent information pertaining to that evaluation.

References

- *Denotes the study was included in the systematic review.
- *Adva. (2008). *Adva against domestic violence and abuse in Devon: The repair program*. Unpublished report. Devon, UK: Author.
- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). Newark, NJ: LexisNexis.
- *Association for the Prevention and Handling of Violence in the Family. (2009). *Love without hurt*. Unpublished report. Cyprus: Author.
- Babcock, J. C., Green, C. E., & Robie, C. (2004). Does batterers’ treatment work? A meta-analytic review of domestic violence treatment. *Clinical Psychology Review, 23*, 1023-1053.
- Barnish, M. (2004). *Domestic violence: A literature review*. London, England: HM Inspectorate of Probation.
- *Bowen, E. (2004). *Evaluation of a community based domestic violence offender rehabilitation program* (Unpublished doctoral dissertation). Centre for Forensic and Family Psychology, University of Birmingham, UK.
- Cavanaugh, M. M., & Gelles, R. J. (2005). The utility of male domestic violence offender typologies: New directions for research, policy, and practice. *Journal of Interpersonal Violence, 20*, 155-166.
- Daly, J. E., & Pelowski, S. (2000). Predictors of dropout among men who batter: A review of studies with implications for research and practice. *Violence and Victims, 15*, 137-160.
- Davis, R. C., & Taylor, B. G. (1999). Does batterer treatment reduce violence? A synthesis of the literature. *Women & Criminal Justice, 10*, 69-93.
- Day, A., Chung, D., O’Leary, P., & Carson, E. (2009). Programs for men who perpetrate domestic violence: An examination of the issues underlying the effectiveness of intervention programs. *Journal of Family Violence, 24*, 203-212.
- Dobash, R. E., & Dobash, R. P. (2000). Evaluating criminal justice interventions for domestic violence. *Crime & Delinquency, 46*, 252-270.
- *Dobash, R. P., Dobash, R. E., Cavanaugh, K., & Lewis, R. (1999). A research evaluation of British programs for violent men. *Journal of Social Policy, 28*, 205-233.
- Dutton, D. G., & Corvo, K. (2007). The Duluth model: A data-impervious paradigm and a failed strategy. *Aggression and Violent Behavior, 12*, 658-667.
- *Echauri Tijeras, J. (2010). *Eficacia de un programa de tratamiento psicológico para hombres maltratadores en la relación de pareja: Características psicopatológicas y resultados terapéuticos* [Efficacy of psychological treatment for domestic violent men: Psychopathological characteristics and therapeutic results] (Unpublished doctoral dissertation). Departamento de Psicología y Pedagogía, Universidad Pública de Navarra, Pamplona-Iruña, Spain.

- *Echeburúa, E., & Fernández-Montalvo, J. (1997). Tratamiento cognitivo-conductual de hombres violentos en el hogar: Un estudio piloto [Cognitive behavioral therapy for violent men: A pilot study]. *Análisis y Modificación de Conducta*, 23, 355-384.
- *Echeburúa, E., & Fernández-Montalvo, J. (2009). Evaluación de un programa de tratamiento en prisión de hombres condenados por violencia grave contra la pareja [Evaluation of a treatment program in prisons for men sentenced for committing domestic violence]. *International Journal of Clinical and Health Psychology*, 9, 5-20.
- *Echeburúa, E., Fernández-Montalvo, J., & Amor, P. (2006). Psychological treatment of men convicted of gender violence: A pilot study in Spanish prisons. *International Journal of Offender Therapy and Comparative Criminology*, 50, 57-70.
- Feder, L., Wilson, D. B., & Austin, S. (2008). Court-mandated interventions for individuals convicted of domestic violence. *Campbell Systematic Reviews*, 12, 1-46.
- Gondolf, E. (2002). *Batterer intervention systems: Issues, outcomes, and recommendations*. Thousand Oaks, CA: Sage.
- Gondolf, E. (2007). Theoretical and research support for the Duluth model: A reply to Dutton and Corvo. *Aggression and Violent Behavior*, 12, 644-657.
- Gracia, E., & Herrero, J. (2006). Acceptability of domestic violence against women in the European Union: A multilevel analysis. *Journal of Epidemiology & Community Health*, 60, 123-129.
- Graham-Kevan, N. (2007). Domestic violence: Research and implications for batterer programs in Europe. *European Journal on Criminal Policy and Research*, 13, 213-225.
- *Hagemann-White, C., Kavemann, B., & Beckmann, S. (2004). Täterarbeit im Kontext von Interventionsprojekten gegen häusliche Gewalt: Wissenschaftliche Begleitung der Interventionsprojekte gegen häusliche Gewalt (WiBiG), Abschlussbericht 2000 bis 2004 [Working with offenders in the context of intervention projects against Domestic Violence: Accompanying research of the intervention projects against domestic violence. Final report: 2000-2004]. In S. Beckmann & C. Hagemann-White (Eds.), *Gemeinsam gegen häusliche Gewalt: Kooperation, Intervention, Begleitforschung* (pp. 691-834). Berlin, Germany: Bundesministerium für Familie, Senioren, Frauen, und Jugend.
- Hamberger, K. L., & Hastings, J. E. (1993). Court-mandated treatment of men who assault their partner: Issues, controversies, outcomes. In N. Z. Hilton (Ed.), *Legal responses to wife assault: Current trends and evaluation* (pp. 188-229). London, England: Sage.
- Hamilton, L., Koehler, J. A., & Lösel, F. A. (2013). Domestic violence perpetrator programs in Europe, Part I: A survey of current practice. *International Journal of Offender Therapy and Comparative Criminology*, 57, 1189-1205.
- Hatcher, R., Palmer, E., Clark, C., Hollin, C., McGuire, J., & Bilby, C. (2003). *Preliminary outcome evaluation of the crime reduction programme pathfinder project* (offending behavior): *Duluth domestic violence programme* (Report prepared for submission to the Correctional Services Accreditation Panel). Unpublished report. UK: Leicester-Liverpool Evaluation Group.
- Koehler, J., Lösel, F. A., Akoensi, T. D., & Humphreys, D. K. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. *Journal of Experimental Criminology*, 9, 19-43.

- *Leicester–Liverpool Evaluation Group. (2005). *Duluth domestic violence reconviction study*. Unpublished Manuscript. UK: Author.
- Lipsey, M., & Cullen, F. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Science*, 3, 297-320.
- Lösel, F. (2007). Doing evaluation research in criminology: Balancing scientific and practical demands. In R. King & E. Wincup (Eds.), *Doing research on crime and justice* (2nd ed., pp. 141-170). New York, NY: Oxford University Press.
- Lösel, F. (2012). Offender treatment and rehabilitation: What works? In M. Maguire, R. Morgan, & R. Reiner (Eds.), *The Oxford handbook of criminology* (5th ed., pp. 986-1016). Oxford, UK: Oxford University Press.
- Lösel, F., Koehler, J. A., Hamilton, L., Humphreys, D. K., & Akoensi, T. D. (2011). Strengthening transnational approaches to reducing reoffending: Final report submitted to the European Commission. Retrieved from http://www.cepprobation.org/uploaded_files/Rep%20STARR%20ENG.pdf
- MacKenzie, D. L. (2006). *What works in corrections: Reducing the criminal activities of offenders and delinquents*. New York, NY: Cambridge University Press.
- Mears, D. (2003). Research and interventions to reduce domestic violence revictimization. *Trauma, Violence, & Abuse*, 4, 127-147.
- Rosenbaum, A. (1988). Methodological issues in marital violence research. *Journal of Family Violence*, 3, 91-104.
- Rosenfeld, B. D. (1992). Court-ordered treatment of spouse abuse. *Clinical Psychology Review*, 12, 205-266.
- Rothman, E. F., Butchart, A., & Cerdá, M. (2003). *Intervening with perpetrators of intimate partner violence: A global perspective*. Geneva, Switzerland: World Health Organization.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Sherman, L., Farrington, D., Welsh, B., & Mackenzie, D. (Eds.). (2002). *Evidence-based crime prevention*. New York, NY: Routledge.
- *Socialstyrelsen. (2010). *Behandling av män som utövar våld i nära relationer - en utvärdering* [Treatment of men who use violence in intimate relationships: An evaluation]. Stockholm, Sweden: Author.
- Stover, C. S. (2005). Domestic violence research: What have we learned and where do we go from here? *Journal of Interpersonal Violence*, 20, 448-454.
- Straus, M. (1991). Conceptualization and measurement of battering: Implications for public policy. In M. Steinman (Ed.), *Woman battering: Policy responses* (pp. 19-47). Cincinnati, OH: Anderson.
- Sullivan, C., Rumpitz, M., Campbell, R., Eby, K., & Davidson, W. (1996). Retaining participants in longitudinal community research. *Journal of Applied Behavioral Science*, 32, 262-276.
- *Törmä, S., & Tuokkola, K. (2009). *Jussi-työ: Miesten perhe-ja lähisuhdeväkivaltatyön ulkoinen arviointi* [Jussi-work: Men in family and domestic violence work, the external evaluation]. Unpublished report. Finland.

- Westmarland, N., & Kelly, L. (2012). Why extending measurements of “success” in domestic violence perpetrator programmes matters for social work. *British Journal of Social Work*. doi:10.1093/bjsw/bcs049
- Wilson, D. (2009). Missing a critical piece of the pie: Simple document search strategies inadequate for systematic reviews. *Journal of Experimental Criminology*, 5, 429-440.