Psychosocial Intervention

Are Interventions with Batterers Effective? A Meta-analytical Review

Ramón Arce, Esther Arias, Mercedes Novo, and Francisca Fariña

Universidad de Santiago de Compostela, Spain

ABSTRACT

The inconsistency in the results both internally and between of previous meta-analyses on batterer intervention program efficacy, and the publication of new batterer interventions underscored the need for an up-to-date meta-analytical review. A total of 25 primary studies were found from literature search, obtaining 62 effect sizes, and a total sample of 20,860 intervened batterers. The results of a global meta-analysis showed a positive, significant, and of a medium magnitude effect size for batterer interventions, but not generalizable. Nevertheless, the results exhibited a significantly higher rate of recidivism measured in couple reports (CRs) than in official records (ORs). As a consequence, intervention efficacy measuring in CRs was null, whilst in ORs was positive and significant. As for the intervention model, positive and significant effects were observed under the Duluth Model and cognitive-behavioural treatment programs (CBTPs), but a higher effect size was obtained with CBTPs in comparison to the Duluth Model (under this model, interventions may have negative effects, i.e., an increase in recidivism rate). In relation to intervention length, short interventions failed to reduce recidivism in ORs and may have negative effects, while long interventions were effective in reducing recidivism rate in ORs without negative effects. Efficacy evaluations in short follow-ups were invalid as artificially boosted recidivism reduction rate. Limitations of ORs and short follow-ups as measures of the intervention efficacy and implications of results for batterer intervention are discussed.

¿Son efectivas las intervenciones con los maltratadores? Una revisión metaanalítica

RESUMEN

La inconsistencia interna y entre las revisiones metaanalíticas en los resultados sobre la eficacia de los programas de intervención con maltratadores, así como la publicación de nuevos estudios, pone de manifiesto la necesidad de llevar a cabo una revisión metaanalítica actualizada. Se encontró un total de 25 estudios primarios, de los que se obtuvieron 62 tamaños del efecto para una muestra total de 20,860 maltratadores intervenidos. Los resultados del metaanálisis global mostraron un tamaño del efecto promedio positivo, significativo y de una magnitud moderada para la intervención con maltratadores, pero no generalizable. Sin embargo, los resultados revelaron una tasa de reincidencia mayor medida en los informes de las parejas (IPs) que en los registros oficiales (ROs). Como consecuencia, la eficacia de la intervención medida en los IPs resultó nula, mientras que en los ROs fue positiva y significativa. En relación al modelo de intervención, se encontraron tamaños del efecto positivos y significativos con el Modelo Duluth y los programas de tratamiento cognitivo-conductuales (PTC-Cs), pero el tamaño del efecto obtenido con los PTC-Cs era significativamente mayor que con el Modelo Duluth (con este modelo las intervenciones pueden tener efectos negativos, es decir, un incremento en la tasa de reincidencia). En relación a la longitud de la intervención, las intervenciones breves fallaron en la reducción de la reincidencia en los ROs y pueden tener efectos negativos, en tanto que las intervenciones largas fueron eficaces en la reducción de la tasa de reincidencia en los ROs y no dan lugar a efectos negativos. Las evaluaciones de la eficacia de la intervención en períodos cortos de seguimiento resultaron no válidas al incrementar artificialmente la tasa de reducción de la reincidencia. Se discuten las limitaciones de la medida de la eficacia de la intervención en los ROs y en períodos cortos de seguimiento, así como las implicaciones para la intervención con maltratadores.

Intervention programs for batterers have been the subject of controversy ever since their conception. These interventions have been open to criticism from both a restorative perspective and a feminist perspective demanding resources should be allocated to victims, not to batterers. Such criticism, however, comes into direct conflict with the legal and judicial mandate of prison institutions that are obliged

to embark on the resocialization and rehabilitation of inmates. A further bone of contention concerns the efficacy of interventions, that is, if the intervention lacks efficacy, it is unfounded. Initially, interventions were designed for violent physical or sexual offenders, but in recent years they have been generalized to less serious offences, mainly minor offences and misdemeanours with non-custodial community sentencing. Claims that interventions with batterers were not effective (Rosenfeld, 1992) have been refuted by meta-analytical reviews that have found interventions to be effective. Nevertheless, in several reviews the mean effect size was significant, but modest (Babcock et al., 2004), whereas in others the mean effect size was important, but not generalizable to all interventions (Arias et al., 2013). Notwithstanding, estimates on the benefits of batterer interventions range from 5% (Babcock et al., 2004) to 20% (Arias et al., 2013). Moreover, the comparison of the mean effect size of the efficacy of interventions (BESD; Rosnow & Rosenthal, 1988) in the most recent meta-analysis (Arias et al., 2013), $r = .20$, with the results for the treatment of delinquency in general (Redondo et al., 2002), and the results for sexual offenders in particular, $r = .13$ (Schmucker & Lösel, 2015), showed batterer interventions were similar or more effective. Moderators of the intervention were analysed as the effects were subject to heterogeneity. The variable measuring effects has been inconsistent, in comparison to the systematic positive and significant, but not generalizable effect reported in official records (ORs) (Arias et al., 2013; Babcock et al., 2004; Feder & Wilson, 2005; Rosenfeld, 1992), the results for couple reports (CRs) were inconsistent. In fact, Feder and Wilson (2005) and Arias et al. (2013) found a null effect in CRs, whilst Rosenfeld (1992) encountered a notable decline in the recidivism rate (33% in treated batterers vs. 47.3% in dropouts), and Babcock et al. (2004) the same effect as in ORs, $d = 0.18$. Likewise, the discrepancies observed in the results for the type of intervention. Thus, in terms of intervention programs based on the Duluth Model, a feminist psychoeducational approach, Babcock et al. observed a significant effect on the reduction of recidivism in ORs, $d = 0.25$ (12.4%) and CRs, $d = 0.24$ (11.9%), but no significant effect was observed in Cognitive-Behavioural Treatment Programs (CBTPs). Moreover, Arias et al. found a moderate effect size, but not generalizable to all studies (i.e., with adverse-negative-outcomes) in ORs both in the Duluth Model, $d = 0.41$, and in CBTPs, $d = 0.47$, which was insignificant in CRs with either type of treatment.

As for the length of the intervention, Babcock et al. (2004) determined short and long interventions were as equally effective in ORs, ($d = 0.16$ and $d = 0.20$, respectively), and Arias et al. (2013) noticed long interventions had a significant and generalizable effect, $d = 0.49$, but short ones did not. In CRs, Babcock et al. obtained small significant effects in both short and long interventions. In contrast, Arias et al. attained both short and long interventions could increase recidivism. During long-term follow-up, Babcock et al. observed a significant effect in the decline of recidivism in ROs, $d = 0.25$, which Arias et al. claimed was not generalizable, and could potentially increase the recidivism rate by 17.7%. In short-term follow-ups, the meta-analysis by Babcock et al. revealed (unexpectedly given that the detected rate of recidivism is obviously smaller in shorter than in longer periods) a smaller effect of the intervention, $d = 0.13$, but significant, than in the work by Arias et al., who obtained a null effect, $d = 0.04$, which could eventually have an adverse effect in the range of 22%. In CRs, Babcock et al. obtained a similar pattern of results: significant effects in both short and long follow-ups, and paradoxically large in long follow-ups that provide a broader measure of recidivism than short periods. In comparison, Arias et al. found a null effect in short follow-ups, $d = 0.03$, with a negative effect of up to 28.3%, and negligible, $d = 0.12$, with adverse effects of up to 18.2% in long follow-ups.

The inconsistency in the results both internally and between meta-analysis, and the fact that initially they measured physical violence, which was recently extended to encompass psychological violence, as well as the proliferation of new batterer intervention programs underscore the need for a meta-analytical review to establish the actual state-of-the-art of batterer interventions, and to elucidate the inter- and intra-analysis inconsistencies.

Method

Study Search

The search of studies was designed to update the 2012 meta-analysis by Arias et al. (2013). Thus, the following search strategies were employed: a) search in broad databases PsycInfo, ERIC, EBSCO, and Google Scholar; b) search in gender violence observatories (v.gr., www.work-with-perpetrators.eu; www.VAWnet.org; www.mincava.umn.edu; www.courtinnovation.org; www.cienciaspenales.net; www.iresweb.org); c) contacting prominent researchers in the field (that is, researchers who were the corresponding authors of all the papers found, both the included and excluded papers); and d) review of all the bibliography in the references lists of all the papers found, and previous meta-analytical reviews. The most productive keywords were: "batterer", "intervention program", "evaluation", "assessment", "effectiveness", "intimate partner violence", "partner-violent men", "recidivism", "reoffending", "attrition", "domestic violence", "court mandates batterer intervention", and "prison intervention".

The relation of keywords was generated by a system of successive approximations whereby the initial keywords of previous meta-analyses determined the inclusion of relevant keywords from all the papers found.

Inclusion and Exclusion Criteria

Bearing in mind the objectives of the meta-analysis, the following inclusion criteria were applied to the papers eligible in the search: a) they reported the number of participants; b) they provided the recidivism rate of the sample of subjects who completed the intervention; c) they applied an experimental or quasi-experimental design (with or without control group); d) they indicated theoretical approach, contents, and duration of the intervention program; and e) they stipulated the follow-up period for measuring recidivism. Studies involving a 6-month follow-up period or shorter were excluded owing to the lack of validity in the measures. A total of 25 papers were included under these criteria, with 62 effect sizes, and a sample total of 20,860 intervened batterers. The study search flowchart is shown in Figure 1.

Data Analysis

A meta-analysis of experiments was performed by correcting the effect sizes according to the distribution of artifacts (Schmidt & Hunter, 2015). As the measure of recidivism was normally expressed in percentages, or when this was not the case they could be computed, in this analysis the measure of recidivism was defined as the percentage of batterers who had reoffended in gender violence (data on recidivism in other offences were excluded) during the follow-up period. Two indices have been proposed for calculating the effect size in proportions: Cohen's h and Hedges and Olkin's (1985) δ, leading to slightly higher, but qualitatively similar results to h (Arias et al., 2013). In this meta-analysis, Hedges and Olkin's δ was employed. Nonetheless, all of the analyses with h were replicated and obtained qualitatively similar results. In terms of interpretation, δ and h are interchangeable with Cohen's d, and an effect size of 0.20, 0.50, 0.80 is considered to be small, medium, and large, respectively. However, this classification has generated discrepant interpretations, where small effect sizes were classified as negligible by certain authors but relevant by others. Hence, the effect size was transformed into percentiles (Monteiro et al., 2018) and the magnitude was interpreted in terms of percentage superiority over all possibilities (Vilariño et al., 2018).
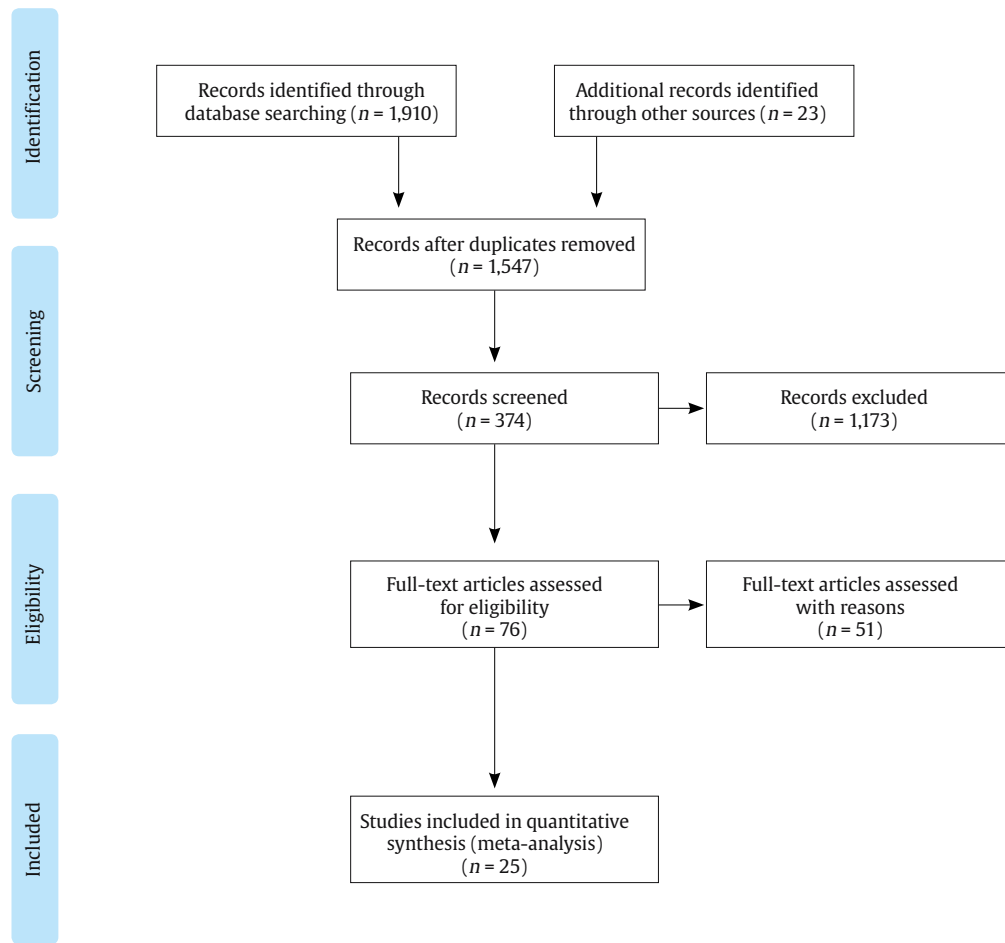
Figure 1. Flow Diagram of the Meta-analysis.

As for the effects of the calculus of δ, in previous reviews it has been calculated with control groups both in experimental and quasi-experimental designs. The results showed the type of design mediated differences in effect sizes. However, the type of analysis design is not an effect of the intervention. That is, the differences in results were due to the test value, not the intervention. For this reason, and in order to homogenize the inter-studies contrast value, to control the bias of control groups in primary studies and to include interventions without control groups, the contrast value of recidivism was calculated as reoffending without intervention and the recidivism rate weighted by the sample error of the control group total (Schmidt & Hunter, 2015), which was .22 in ORs and .28 in CRs. Computed effect sizes from primary studies (see Appendix), the following statistics were calculated: sample size weighted mean effect size (d), standard deviation of d (SD_d), standard deviation predicted for sampling error alone (SD_{pre}), standard deviation of d after removing sampling error variance (SD_{res}), mean true effect size (δ), standard deviation of δ (SD_δ), percentage of variance attributable to statistical artifacts (%Var), 95% confidence interval for d (95% CI_d), and 80% credibility interval for δ (80% CI_δ).

Though results opened the door to further meta-analyses (the variance explained by artifactual errors was below 75%, i.e., 75% rule; Schmidt & Hunter, 1981), further meta-analyses were impossible as number of sizes ($k ≤ 3$) or sample sizes ($N < 400$) were insufficient, or no other moderators had been defined in the primary studies.

When meta-analyses are calculated for the study of moderators for the grouping of variables according to levels, the results do not include the comparison between meta-analyses. Thus, the results were insufficient, as is the case of the analysis of intervention models where significant effects were observed in ORs, but if they were the same or higher in one condition or another was not reported. The same occurred with the duration of the intervention, and the follow-up period. To overcome this contingency, the solution proposed by Amado et al. (2015) was adopted, where the statistic q_s was calculated to compare two effect sizes by converting r, and then comparing the r_s. Furthermore, the results of the meta-analysis were also quite limited in discerning the implications for practice, in this case, the intervention. Regarding the implications of the meta-analysis on professional practice, Fariña et al. (2017) recommend, according to the specific objectives at hand, sensitivity estimation (statistical U), effect quantifying (in present meta-analytical analysis, the efficacy of the intervention, BESD), and the probability of superiority.

Criterion Reliability

As CRs were gathered with different measurement instruments, the correction for the criterion unreliability was computed with Mosier's (1943) composite reliability coefficient, $r = .87$.

Coding

For the analysis of moderators, the following recidivism variables were encoded (ORs, $N = 19,429$ and $k = 46$, and CRs, $N = 1,351$ and $k = 16$); follow-up time ($≤ 12$ months, $N = 3,509$ and $k = 21$, and > 12

Table 1. Meta-Analyses of the Effectiveness of Batterer Intervention Programs

k	N	d	SD_d	SD_{pre}	SD_{res}	δ	$SD_δ$	% Var	95% CI_d	80% $CI_δ$
General Meta-Analysis										
[1]62	20,860	0.43	0.4485	0.1114	0.4345	0.44	0.4300	6.17	0.40, 0.46	-0.13, 1.00
[2]59	20,215	0.48	0.3278	0.1108	0.3275	0.49	0.3332	10.28	0.45, 0.51	0.06, 0.91
Meta-Analysis of Intervention Effectiveness Measured in CRs										
[1]16	1,351	0.004	0.2218	0.2194	0.0321	0.005	0.0344	97.90	-0.10, 0.11	-0.03, 0.05
[2]15	1,340	-0.008	0.0000	0.2131	0.0000	-0.009	0.0000	100.00	-0.11, 0.09	-0.009
Meta-Analysis of Intervention Effectiveness Measured in ORs										
[1]46	19,509	0.45	0.4490	0.0988	0.4338	0.45	0.4338	4.93	0.42, 0.48	-0.10, 1.01
[2]44	18,875	0.51	0.3279	0.0984	0.3129	0.51	0.31298	99.00	0.48, 0.54	0.11, 0.91
Meta-Analysis of Intervention Effectiveness Measured in ORs on Short Follow-up Periods										
[1]21	3,509	0.35	0.8191	0.1563	0.8041	0.35	0.8041	3.64	0.28, 0.42	-0.68, 1.37
[2]19	2,875	0.69	0.4004	0.1679	0.3635	0.69	0.3635	17.58	0.61, 0.77	0.23, 1.15
Meta-Analysis of Intervention Effectiveness Measured in ORs on Long Follow-up Periods										
26	16,050	0.48	0.3026	0.0818	0.2914	0.48	0.2914	7.31	0.45, 0.51	0.11, 0.85
Meta-Analysis of Intervention Effectiveness of Duluth Model Measured in ORs										
[1]25	15,027	0.37	0.4627	0.0824	0.4553	0.37	0.4553	3.17	0.34, 0.40	-0.20, 0.96
[2]23	14,393	0.44	0.3168	0.0811	0.3168	0.44	0.3168	6.15	0.41, 0.47	0.04, 0.85
Meta-Analysis of Effectiveness of CBTPs Measured in ORs										
9	1,629	0.88	0.1382	0.1562	0.0000	0.88	0.0000	100.00	0.78, 0.98	0.88
Meta-Analysis of Effectiveness of OTIs Measured in ORs										
12	2,853	0.63	0.2155	0.1333	0.1693	0.63	0.1693	38.26	0.55, 0.71	0.41, 0.85
Meta-Analysis of Effectiveness of Brief Interventions Measured in ORs										
[1]14	3,631	0.03	0.7064	0.1244	0.6953	0.03	0.6953	3.10	-0.03, 0.09	-0.86, 0.92
[2]12	2,997	0.29	0.4379	0.1274	0.4379	0.29	0.4379	7.81	0.22, 0.36	-0.27, 0.85
Meta-Analysis of Effectiveness of Long Interventions Measured in ORs										
32	15,878	0.55	0.2785	0.0917	0.2630	0.55	0.2630	10.84	0.52, 0.58	0.21, 0.89

Note. k = number of effect sizes; N = total sample size; d = sample size weighted mean effect size; SD_d = standard deviation of d; SD_{pre} = standard deviation predicted for sampling error alone; SD_{res} = standard deviation of d after removing sampling error variance; δ = mean true effect size; $SD_δ$ = the standard deviation of δ; %var = percent of observed variance accounted by artifactual errors; 95% CI_d = 95% confidence interval for d; 80% $CI_δ$ = 80% credibility interval for δ; [1]meta-analysis including all effect sizes from primary studies; [2]meta-analysis removing extreme and/or outliers effect sizes.

months, N = 16,050 and k = 26); treatment duration (< 16 sessions/weeks, N = 3,631 and k = 14; and >16 sessions/weeks, N = 15,878 and k = 32); intervention level (individual vs. multi-level); type of sessions (individual sessions, group sessions or combined sessions); and type of treatment (Duluth Model: N = 15,027 and k = 25; CBTPs: N = 1,629 and k = 9, and OTIs: N = 2,853 and k = 12).

The encoding was carried out by an encoder who noted the levels of each of the categories created and described by researchers, and marked exactly where they were referred to in the text. In the encoding of the type of intervention, the criterion stated by the authors of the primary studies was applied, but it should be noted that increasingly interventions based on the Duluth Model also include cognitive-behavioural training techniques for anger management, and CBTPs a gender perspective (e.g., patriarchal attitudes and values). Notwithstanding, the approach ascribed to the intervention was in line with the models proposed. A second encoder reviewed all of the studies using the same encoding system. Thereafter, both the registered categories, and the exact correspondence in the encodings were checked to estimate the true concordance ($\overline{k}$) with the kappa statistic (Arce et al., 2000). Kappa corrects the random effect in concordance, but not the correspondence in encodings. Thus, the registration of a category by two encoders in two different places are two encoding errors in the true kappa (lack of correspondence), whereas it is encoded as correct with Cohen's kappa. Moreover, consistency between two encoders is not sufficient for estimating the fidelity of encodings (i.e., coding correctness in relation to the content of the category), which would require for the encoding to remain stable through time, intra-encoder, and in other evaluation contexts (coders in other studies). The true concordance inter- and intra-encoders was very good ($\overline{k}$ > .81). Moreover, the encoders were consistent in other contexts

i.e., evaluations (Arias et al., 2013; Gallego et al., 2019). Bearing in mind the inter- and intra-evaluator and inter-context consistency of coders, encodings were reliable, that is, reflected the original data.

Results

Contrast of Inter-criteria Consistency of the Recidivism Measurement

The results showed a significantly higher rate (+.1540), Z = 12.32, p < .001, of recidivism measured in CRs, than in ORs. Thus, ORs masked a significant amount of recidivism (not surprisingly, because many couples refuse to collaborate, a predictable behaviour associated to revictimization; Brame et al., 2015; and concealment of injuries; Arce, Fariña, Seijo, et al., 2015). Moreover, the confidence margin of the mean rate of recidivism, higher in CRs than in ORs, CI 95% [.1491, .1589], revealed it was stable around 15%.

Analysis of Outliers

Taking into account that the efficacy of a treatment varies according to the variable measured, an analysis of outliers was carried out for each measure, ORs and CRs. Data showed that one of the studies, Stith et al. (2004) with CRs, was an extreme value (± 3 * IQR). As for the effect sizes of studies measuring recidivism in ORs, no extreme cases were observed nor outliers with the criterion ± 1.5 * IQR, but extreme cases were observed with a much more conservative criterion ± 2SD (Chauvenet's criterion). This criterion requires that each meta-analysis should verify the correct classification of outliers and inconvenient studies (contrary

to the hypothesis of analysis), or moderators (Tukey, 1960). Thus, a meta-analysis was performed with and without these studies to determine their impact and their effects on estimators. Moreover, a meta-analysis with and without extreme values was performed.

Analysis of the General Efficacy of Batterer Intervention Programs

The meta-analysis of the total effect size of the studies on the efficacy of batterer interventions (see Table 1) showed, for 62 effect sizes and for a sample of 20,860 batterers, a positive, significant (confidence interval for d does not include zero), and medium-moderate mean true effect size, $\delta = 0.44$, for batterer interventions. However, these results were not generalizable (credibility interval for δ includes zero) to all studies on batterer interventions, since interventions may have negative effects. Likewise, the results without outliers ($N = 20,215$ and $k = 59$) showed a positive, significant (confidence interval for d does not include zero), medium magnitude, and generalizable (credibility interval for δ does not include zero), i.e., batterer interventions had no negative effects (on recidivism) mean true effect size, $\delta \approx 0.50$. Comparatively, the effect sizes with and without atypical values were equivalent, $q_s = .024$, *ns*. Thus, the elimination of outliers excluded inconvenient results that had no effects on global results. In terms of net intervention, efficacy (reduction in recidivism rate) would be approximately 21.49% ($r = .2149$), but with a negative lower limit, i.e., an increase in recidivism of 6.49% (80% LCV [lower credibility value] converted to $r = -.0649$). The magnitude of the effect size was greater than 62.17% of all possibilities, and 24.34% of all the positive effect sizes of the interventions.

Nevertheless, the percentage of variance explained by artifactual errors was below 75% in both meta-analyses, so these results were mediated by moderators of effect.

The most extensively researched moderator is the variable measuring effects (criterion), i.e., recidivism, which has been assessed in primary studies in both ORs (i.e., police, courts, correctional institutions), and CRs.

Analysis of the Effects of Batterer Interventions on the Variable Measuring Recidivism

The meta-analysis on the studies measuring intervention efficacy on the recidivism rate in CRs (see Table 1), with a sample of 1,351 batterers and 16 effect sizes, revealed that interventions had no effect on recidivism, with a null ($\delta = 0.005$) mean true effect size ($U1 = 0.007$, i.e., the independence of the distributions of treated and non-treated batterers was only 0.7%), and could be negative by up to -0.10 or, in other words, the intervention could have a negative effect increasing recidivism rate by up to 4.99% ($r = -.0499$). The lack of effects of the intervention measured in the recidivism rate in CRs was not mediated by moderators (% VAR > 75). Thus, the results were conclusive. The results were replicated (see Table 1) including the extreme case ($N = 1,340$ and $k = 15$), which underpins their stability.

The meta-analysis on studies estimating intervention efficacy on recidivism in ORs, with a sample of 19,509 batterers and 46 effect sizes, showed (see Table 1) a positive, significant (confidence interval for d does not include zero), small-medium ($\delta = 0.45$) and non-generalizable (credibility interval for δ includes zero) mean true effect size in the intervention, with possible negative effects of up to 4.99% (80% LCV converted to $r = -.0499$), which were mediated by moderators of the relationship between treatment and recidivism (% VAR > 75%). Without outliers ($N = 18,875$ and $k = 44$), the results were replicated: a positive, significant, medium, and mediated by moderators mean true effect size, but generalizable (credibility interval for δ does not include zero) to all the studies. Comparatively, the effect size of the meta-analysis

with and without outliers were similar, $q_s = .029$, *ns*. In short, the effects of outliers were constrained to the lack of generalization of the results to all of the studies, so by eliminating them we would be discarding inconvenient results, not real outliers. Thus, the mean recidivism reduction rate in ORs due to the intervention (versus non-intervened offenders) was 21.95%, but with a lower negative limit, that is, an increase in recidivism of up to 4.99%. Succinctly, on average, interventions reduced recidivism in ORs, but they could also have adverse effects by increasing recidivism rate by more than 50% (Taylor & Maxwell, 2009). The magnitude of the effect was higher than 62.55% of all possibilities, and higher than 25.1% of all the positive effect sizes for batterer interventions. Thus, the positive intervention effect is explained by the measurement method, ROs, rather than the measured construct, recidivism (Podsakoff et al., 2003). However, the percentage of variance explained by artifactual errors was lower than 75%. So, the results were mediated by moderators.

Analysis of the Effects of Follow-up Time on Recidivism in ORs

The recidivism follow-up period is a critical factor for criterion validity. In fact, short follow-up periods can artificially increase the efficacy rate, given that approximately 2/3 of reoffending occurs in the first two years (Redondo et al., 2001). According to Gondolf, 2000 and Jones and Gondolf (2002), reoffending occurs in batterers in half the time, and there is widespread agreement that most battery goes unreported, and not documented in ORs (European Union Agency for Fundamental Rights, 2014). For this reason, the follow-up period was taken as a moderator of the results of the intervention and were subdivided into short follow-up periods of 6 to 12 months, and long follow-up periods of more than 12 months, a classification attested to be valid (Arias et al., 2013; Gondolf, 2000).

The meta-analysis on the efficacy of batterer interventions in short follow-up periods, i.e., 12-month or less in ORs (see Table 1), with a sample of 3,509 batterers and a total of 21 effect sizes, found a positive and significant (confidence interval for d does not include zero) mean true effect size ($\delta = 0.35$), but not generalizable to other studies (credibility interval for δ includes zero), and of a small magnitude. These results were replicated in the meta-analysis without outliers ($N = 2,875$, $k = 19$), with a positive, and significant (confidence interval for d does not include zero) mean true effect size of moderate magnitude ($\delta = 0.69$), which was generalizable to other studies (credibility interval for δ does not include zero). Comparatively, the effect size of the intervention in the meta-analysis without outliers was significantly higher than with outliers, $q_s = .164$, $p < .01$, which was an unexpected and incongruous result (outliers diminished the efficacy of the intervention). Therefore, the results should be analysed without outliers. In practical terms, batterer interventions reduced the recidivism rate in ORs by an average 32.61%, without interventions with adverse effects, (80% LCV = .23), and the effect size was higher than 68.79% of all possibilities and 37.58% of all positive intervention effect sizes. However, these results were mediated by moderators, both in the meta-analysis with all of the effect sizes, and without outliers (% VAR < 75).

The meta-analysis on the efficacy of batterer interventions on 12-month follow-up periods longer in ORs (see Table 1), with 26 effect sizes and a sample of 16,050 batterers, revealed a positive and significant (confidence interval for d does not include zero) mean true effect size of moderate magnitude ($\delta \approx 0.50$). Moreover, this result was generalizable to other studies (credibility interval for δ does not include zero), but was mediated by moderators (% VAR < 75). As for the reduction in the recidivism rate in ORs, batterer interventions reduced the recidivism rate by 23.34% ($r = .2334$) and, once again, without detecting negative intervention effects (80% LCV = .11). The

magnitude of the effect size was above the 63.31% of all possibilities and 26.62% of all the positive intervention effect sizes.

The mean true effect size of the intervention in ORs with a short follow-up period of up to 12 months ($\delta = 0.69$) or less was significantly larger, $q_s = .101$, $p < .05$, than that obtained for the long follow-up period of more than 12 months ($\delta = 0.48$).

Moderators could not be analysed as Ns (< 400) or effect sizes were insufficient ($k \leq 3$).

Analysis of the Effects of the Modality of the Intervention on Recidivism in ORs

Another moderator reported in the literature as having effects on the intervention is the modality of the intervention, categorized as interventions based on the Duluth Model, Cognitive-Behavioural Treatment interventions, and Other Type of Interventions (Arias et al., 2013; Babcock et al. 2004; Feder & Wilson, 2005; Levesque & Gelles, 1998).

Meta-analysis of the effects on recidivism in ORs for interventions based on the Duluth Model with a sample of 15,027 batterers, and 25 effect sizes revealed (see Table 1) a positive and significant (confidence interval for d does not include zero) mean true effect size of small magnitude ($d = 0.37$). Moreover, these results were not generalizable (credibility interval for δ includes zero) to all studies with interventions based on the Duluth Model. These results were replicated without outliers ($N = 14,393$, $k = 23$), with the exception that they were generalizable to all the studies (credibility interval for δ does not include zero). Comparatively, the effect sizes of all the studies, and those without outliers were similar, $q_s = .034$, *ns.* Thus, the outliers were not such, but inconvenient studies with negative effects. In practical terms, interventions based on the Duluth Model reduced on average recidivism by 18.19% ($r = .1819$), with potential negative effects of up to 9.95% (80% LCV converted to $r = -.0995$), and an effect size above 60.25% of all possibilities, and 20.50% of all the positive intervention effect sizes. However, the results were mediated by moderators (% VAR < 75).

The meta-analysis of the effects of batterer interventions with CBTPs on recidivism in ORs (see Table 1), with a total of 9 effect sizes and a sample of 1,629 batterers, exhibited a positive, significant for the intervention (confidence interval for d does not include zero), and of a large magnitude ($\delta = 0.88$) mean true effect size, implying an average reduction in the recidivism rate of 40.27% ($r = .4027$). The generalization of the true effect size could not be studied as the variance was zero (i.e., the studies of this meta-analysis were not randomly distributed); thus, it was analysed with the effect size weighted by sampling error (conservative estimate versus the true size), and was generalizable (credibility interval for d does not include zero) to other studies, and without negative effects, 80% CI [0.70, 1.06]. Additionally, the lower limit of the effect (90% of the interventions surpassed this limit) was 33.40% (80% LCV converted to $r = .3340$). The magnitude of the effect size was higher than 73.24% of all possibilities, and 46.48% of all the positive intervention effect sizes.

The meta-analysis of the effects of Other Types of Intervention (OTIs) on recidivism in ORs (see Table 1), with a total of 12 effect sizes, and a sample of 2,853 batterers, found a positive, moderate, and significant (confidence interval for d does not include zero) mean true effect size ($\delta = 0.63$), generalizable to all the studies (credibility interval for δ does not include zero). The reduction in the recidivism rate with OTIs (versus the rate for non-intervened batterers) was on average 30.04% ($r = .3004$), and without interventions with negative effects, and a minimum threshold in the reduction of the recidivism rate of 20.08% (80% LCV converted to $r = .2008$). The reduction in the recidivism rate by OTIs (versus the rate for non-intervened batterers) was on average 30.48%, without negative effects, and with a minimum threshold in the reduction of recidivism of 20.55%. The magnitude of

the effect size was higher than 67.36% of all possibilities, and 34.72% of all the positive intervention effect sizes. Nevertheless, the effect of these intervention modalities on recidivism was mediated by moderators (% VAR < 75). Once again, the analysis of moderators in the different types of interventions was not possible due to insufficient Ns and ks.

Comparatively, the effect of the intervention on the reduction of recidivism was significantly higher in the CBTPs than in the Duluth Model, $q_s = .243$, $p < .01$ and in the OTIs, $q_s = .117$, $p < .01$, and in the OTIs than in the Duluth Model, $q_s = .126$, $p < .01$.

Analysis of Effects of Duration of the Intervention on Recidivism in ORs

A further moderator of the effects that is systematically reviewed concerns the duration of the intervention (Arias et al., 2013; Babcock et al., 2004). In accordance with Babcock et al. (2004), two categories were created: short duration for programs under 16 sessions/weeks (4 months) and long duration for programs of more than 16 weeks/ sessions.

The results of the meta-analysis for short interventions (< 16 sessions), with a total of 14 effect sizes and 3,631 batterers (see Table 1), showed a non-significant (confidence interval for d includes zero) mean true effect size and, consequently, it was not generalizable to other studies. The meta-analysis replicated without outliers ($N = 2,997$, $k = 12$) found a positive and significant (credibility interval for δ does not include zero) mean true effect size of small magnitude ($\delta = 0.29$), not generalizable (credibility interval for δ includes zero), and mediated by moderators (% VAR < 75) that could not be analysed owing to the insufficient Ns and ks. The comparison of effect sizes with all of the studies with the exclusion of outliers revealed significant differences, $q_s = .129$, $p < .01$, ranging from a non-significant (all of the studies) to a significant (without outliers) mean true effect size. Thus, the outliers were not such, but inconvenient studies with very negative effects. In fact, brief programs can increase the recidivism rate by 39.5% (80% LCV converted to $r = .3950$).

Regarding long intervention programs (> 16 sessions), the results of the meta-analysis with a total of 32 effect sizes and a sample of 15,878 batterers (see Table 1) displayed a positive and significant (confidence interval for d does not include zero) mean true effect size of moderate magnitude ($\delta = 0.55$), generalizable to other studies (credibility interval for δ does not include zero). As for the efficacy of the intervention, long interventions reduced the recidivism rate in ORs by 26.52% ($r = .2652$), and without interventions with negative effects, and a minimum threshold in the reduction of recidivism of 10.44% (80% LCV converted to $r = .1044$). The magnitude of the effect size was higher than 65.17% of all possibilities and 30.34% of all the positive intervention effect sizes. However, the average effect size was mediated by moderators (% VAR < 75) that were not studied owing to insufficient Ns and ks.

Comparatively, long programs were more effective in reducing recidivism in comparison to short interventions, $q_s = .257$, $p < .01$.

Discussion

The present meta-analyses have several limitations that should be borne in mind in the generalization of results. First, the classification of studies taken from primary studies outlined the characteristics for defining them (e.g., an intervention was categorized under a particular model, but the description also encompassed other models; the length of the follow-up was differently measured); thus, it was categorically assumed in the meta-analysis, when there may be substantial variability in categories or overlapping categories. Second, not all studies consistently reported moderators (e.g., treatment type), with the subsequent exclusion of them. Third, the validity of the measure of

efficacy of interventions on recidivism failed to be a good estimator, as it may not even be the real objective of interventions under Spanish law treatment in correctional institutions is voluntary, and the mandate of correctional institutions is to develop aptitudes to overcome deficiencies and to modify unfavourable and negative attitudes; Ley Orgánica 1/1979, de 26 de septiembre, General Penitenciaria). Fourth, the evaluation of the efficacy of the intervention on recidivism based on official records was a criterion of limited validity (a strong positive biased approximation of recidivism; in other words, in comparison to CRs, it failed to capture approximately half of the recidivism). Fifth, follow-up time, as a measure of the efficacy of interventions on recidivism, had a direct and big effect on the validity of the measure. In general, an estimated 50 to 60% of recidivism occurred during the first year (Redondo et al., 2001), with a total confirmed recidivism rate ranging from 40% to more than 80%, which varied according to the measure and the follow-up time (Stover, 2005). Sixth, the measure of the efficacy of interventions in CRs was also probably positively biased (overestimation of the results of the interventions) both in the measure itself (e.g., not all of the batterers who had a partner after treatment reported it; the exposure times of couples with direct effects on recidivism were not reported; a significant number of women failed to respond to encoders) and due to response bias (e.g., bias of leniency, i.e., tendency to minimize or conceal assaults). In fact, one of the distinctive characteristics of victims of gender violence (injury) is the failure to report offences and the concealment of assaults in reports (Holt et al., 2003), concealing offences (Arce, 2017), minimizing injuries (Arce, Fariña, & Vilariño, 2015; Vilariño et al., 2009), and failing to report the recidivism of intervened batterers (Brame et al., 2015), in particular during cohabitation with the batterer. Seventh, evaluations of efficacy were taken from the authors of the interventions, who rely on efficacy for program continuity. Thus, the number of unpublished studies with no effects or negative effects may be high. Eighth, test values were also subject to similar measurement errors. Ninth, the dropouts linked to recidivism were not generally encoded as such, which would artificially increase the efficacy rate (Taft et al., 2001). Tenth, other moderators (not studied as Ns or ks were insufficient) may modify the magnitude or direction of the treatment efficacy (Martín & Moral, 2019; Martín et al., 2019). Thus, further research with additional moderators should be designed. In short, the results had a certain degree of validity for measuring the efficacy of interventions versus non-intervention, but not sufficient to determine the real rates of recidivism, which were presumably significantly higher than registered.

These limitations, in terms of validity of conclusions, can be classified as systematic measurement errors (when they generate an alternative explanation to the results), or a biased measurement method (i.e., variance attributed to method and not to measured construct) (Podsakoff et al., 2003). Bearing in mind the implications of these limitations on the validity of the conclusions of this meta-analysis, they are discussed further below:

a) The control and analysis of the effects in the results of extreme values and outliers was open to two different interpretations. The extreme value referred to a zero recidivism rate, an impossible result that also distorts results. Moreover, outliers (the classification criterion was the statistical significance, with a probability of occurrence of 5%, versus the common criterion ± 1.5 * IQR, much more stringent and that would not classify any as an outlier) were generally (with the exception of short follow-up periods) inconvenient studies (contrary to the hypothesis of the efficacy of interventions postulated in the analysis itself), which were not precisely outliers, but moderators (Tukey, 1960). Therefore, they are valid cases where interventions are observed to have no results or adverse effects, a possibility that has been validated in previous reviews (Arias et al., 2013; Babcock et al., 2004; Feder & Wilson, 2005; Levesque y Gelles, 1998).

b) The recidivism rate was almost 14% in intervened batterers, nearly 13% (.1295) in ORs, and 26.13% (.2613) in CRs. Thus, half of recidivism went unreported in ORs. Nevertheless, this was still far from the true rate, given that, as previously stated, sources of measurement error were inherent to the method itself (measure of intervention efficacy), not the construct under assessment (recidivism), and tended to maximize the efficacy of interventions (variance attributed to the method, not to the construct). Hence, evaluation was a source of systematic error: maximizing the effects of interventions.

c) The results of the general meta-analysis revealed that intervened batterers benefited from interventions compared to non-intervened batterers. According to this meta-analysis, the general efficacy in the reduction of the recidivism rate was approximately 21%. Though the efficacy of interventions was estimated to be as negligible in reviews on the issue (Babcock et al., 2004), this would not be so in relation to the total potential intervention effects, where efficacy obtained would be above 62.17%, $P_{62.17}$, and 24.34% of positives (reduction in recidivism rate). Furthermore, efficacy was higher than that obtained for delinquency in general (Redondo et al., 2002), and higher than that obtained for sexual batterers (Schmucker & Lösel, 2015). Nevertheless, interventions can have no effect on recidivism, or even negative effects (increase recidivism). In short, in general interventions, reduced measured recidivism rates, but not all interventions, were efficacious.

d) Most of the effects of interventions (93.5%) were measured in recidivism in ORs. Meta-analysis on the efficacy of interventions with this measure showed a significant reduction in recidivism rate (a smaller probability of being accused or sentenced for intimate partner violence than non-intervened batterers), estimated to be around 22%, an efficacy rate above 62.55%, $P_{62.55}$, of all possibilities, and 25.1% of positive effects. Once again, not all interventions had the effect of reducing recidivism, and some may even have negative effects (increase recidivism), implying not all interventions were valid.

e) The follow-up period on efficacy of interventions mediated differences in recidivism in ORs. Whilst recidivism rate in short follow-ups fell to around 33%, reduction in long follow-ups was 23%, which was a statistically significant reduction. Thus, short follow-up periods artificially boosted recidivism reduction rate by 10%. Moreover, these results corroborated that 2/3 of all recidivism, in accordance with results by Redondo et al. (2001), occurred during the first two years, or in half that time for batterers according to Gondolf, 2000 and Jones and Gondolf, 2002. Thus, evaluations in short follow-up periods were invalid.

f) Intervention contents had a direct impact on the reduction of batterer recidivism in ORs. As for intervention model, the highest effects were obtained with CBTPs, a result that has been systematically substantiated in literature reviews (Arias et al., 2013; Babcock et al., 2004; Feder & Wilson, 2005; Levesque & Gelles, 1998), and supports the view that interventions should be multimodal: cognition and behaviour (Lila et al., 2019). A small, and significantly lower effect, was obtained with interventions based on the Duluth Model, which could even have adverse effects (more recidivism). The small effect size, in addition to the high number of dropouts during these interventions, has led Taft et al. (2001) to the conclusion that these types of intervention were possibly not effective. This claim was not corroborated by the results, as effect size was significant and associated to a mean recidivism reduction rate of approximately of 18%.

Notwithstanding, the high variability observed could have negative effects. If we take into account that programs based on the Duluth Model were initially aimed at educating in a gender perspective, but were subsequently transformed into cognitive-behavioural models with a gender perspective, addition of gender perspective may be the cause for the decline in efficacy versus exclusively cognitive-behavioural interventions. That is, gender perspective of intervention may fail to re-educate batterers, or even have adverse effects (reactance to treatment or a feigning adherence to treatment could explain this contingency).

g) OTIs (generally focused on cognition and behaviour, but supplemented and guided by other intervention models, such as ecological, therapeutic, or multilevel models) had the effect of significantly reducing recidivism in ORs, and intermediate between Duluth Model and CBTPs. Nevertheless, these results were limited to the meta-analysis studies and were not generalizable to models other than those examined (this category includes any model that was not strictly cognitive-behavioural or Duluth). A close inspection of results indicates cognitive-behavioural interventions in addition to an ecological or multilevel intervention models improved the efficacy of interventions, thereby the need for further studies to evaluate both of these combined models.

h) Short interventions failed to reduce recidivism in ORs and could have negative effects (increase recidivism as compared to no intervention) by almost 40%. In comparison, long interventions were effective in reducing recidivism rate in ORs (≈ 26%), without negative effects. This underscores batterer interventions should be long-term to be effective. Batterers harbour toxic cognitions (Maruna, 2004), which are highly resistant to interventions (Novo et al., 2012), in particular short interventions that may be ineffective or even counterproductive.

i) However, in terms of recidivism in CRs, the efficacy of interventions versus non-intervention was totally null, with a definite result (i.e., generalizable and not mediated by moderators). Hence, interventions were effective in reducing the rate of accusations and guilty verdicts, but not the rate of real, direct intimate partner violence. This result underscored the limitations of ORs as a measure of the efficacy of interventions, that is, ORs were a poor valid measure as a substantial amount of recidivism went undetected. Several causes may be responsible for this limitation in validity: unreported recidivism, rejected complaints by the police, not all gender violence recorded in CRs constitutes gender violence across all countries, and it may be recorded as sexist violence which falls out of the scope of batterer interventions.

In conclusion, there is a corpus of literature on the efficacy of interventions, showing significant effects in reducing recidivism in official records. In other words, intervened batterers were less likely to be accused/sentenced again in (ORs) for the same offence. Notwithstanding, not all of the interventions were effective in ORs. Thus, short interventions were completely ineffective and could have negative effects of up to 40%, and certain interventions based on the Duluth Model may have negative effects of up to 10%. In contrast, long interventions based on CBTPs or OTIs (the results may not be generalised to other techniques than those revised) were on average effective and without negative effects on recidivism in ORs. The evaluation of the efficacy of the intervention in ORs was not a valid measure as it was subject to a systematic measurement error (it minimized the rate), which could reach over 80% (Stover, 2005). However, regardless the fact that ORs were not a valid measure of recidivism, we may draw from the meta-analysis with ORs the conclusion that intervention programs should be based on a long cognitive-behavioural approach.

However, this does not imply that interventions had a significant effect on reducing recidivism in all measures of violence. There was a null effect in couple reports. Further research should be undertaken to evaluate the efficacy of interventions in both types of measure to identify the type of recidivism lost in ORs and to identify moderators of the efficacy of interventions on this measure. Moreover, studies are required to assess the efficacy of interventions on other criteria such as the internal mechanisms underlying gender violence (Arce et al., 2014), and acquisition of skills and abilities that enable batterers to successfully manage risk situations of intimate partner violence (Arce & Fariña, 2010; Lila et al., 2018; Martín-Fernández, Gracia, Marco, et al., 2018; Martín-Fernández, Gracia, & Lila, 2018).

Conflict of Interest

The authors of this article declare no conflict of interest.

References

References marked with an asterisk indicate studies included in the meta-analysis

Amado, B. G., Arce, R., & Herraiz, A. (2015). Psychological injury in victims of child sexual abuse: A meta-analytic review. *Psychosocial Intervention, 24*(1), 49-62. https://doi.org/10.1016/j.psi.2015.03.002

Arce, R. (2017). Análisis de contenido de las declaraciones de testigos: evaluación de la validez científica y judicial de la hipótesis y la prueba forense [Content analysis of the witness statements: Evaluation of the scientific and judicial validity of the hypothesis and the forensic proof]. *Acción Psicológica, 14*(2), 171-190. https://doi.org/10.5944/ap.14.2.21347

Arce, R., & Fariña, F. (2010). Diseño e implementación del programa Galicia de reeducación de maltratadores: una respuesta psicosocial a una necesidad social y penitenciaria [Design and implementation of the Galicia program for batterer re-education: A psychosocial answer to a social and penitentiary need] *Psychosocial Intervention, 19*(2), 153-166. https://doi.org/10.5093/in2010v19n2a7

Arce, R., Fariña, F., & Fraga, A. I. (2000). Género y formación de juicios en un caso de violación [Gender and juror judgment making in a case of rape]. *Psicothema, 12*(4), 623-628. http://www.psicothema.com/pdf/381.pdf

Arce, R., Fariña, F., & Novo, M. (2014). Competencia cognitiva en penados primarios y reincidentes: implicaciones para la reeducación [Cognitive competence among recidivist and non-recidivist prisoners: Implications for the rehabilitation]. *Anales de Psicología, 30*(1). 259-266. https://doi.org/10.6018/analesps.30.1.158201

Arce, R., Fariña, F., Seijo, D., & Novo, M. (2015). Assessing impression management with the MMPI-2 in child custody litigation. *Assessment, 22*(6), 769-777. https://doi.org/10.1177/1073191114558111

Arce, R., Fariña, F., & Vilariño, M. (2015). Daño psicológico en casos de víctimas de violencia de género: un estudio comparativo de las evaluaciones forenses [Psychological injury in intimate partner violence cases: A contrastive analysis of forensic measures]. *Revista Iberoamericana de Psicología y Salud, 6*(2), 72–80. https://doi.org/10.1016/j.rips.2015.04.002

Arias, E., Arce, R., & Vilariño, M. (2013). Batterer intervention programmes: A meta-analytic review of effectiveness. *Psychosocial Intervention, 22*(2), 153-160. https://doi.org/10.5093/in2013a18

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*(8), 1023-1053. https://doi.org/10.1016/j.cpr.2002.07.001

*Babcock, J. C., & Steiner, R. (1999). The relationship between treatment, incarceration and recidivism of battering: A program evaluation of Seattle's coordinated community response to domestic violence. *Journal of Family Psychology, 13*(1), 46-59. https://doi.org/10.1037//0893-3200.13.1.46

*Bennett, L., Call, C., Flett, H., & Stoops, C. (2005). *Program completion, behavioral change and re-arrest for the batterer intervention system of Cook County, Illinois.* Illinois Criminal Justice Information Authority. Chicago, IL. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.589.9366&rep=rep1&type=pdf

*Bowen, E., Gilchrist, E. A., & Beech, A. R. (2005). An examination of the impact of community-based rehabilitation on the offending behaviour of male domestic violence offenders and the characteristics associated with recidivism. Legal and *Criminological Psychology, 10*(2), 189-209. https://doi.org/10.1348/135532505x36778

Brame, R., Kaukinen, C., Gover, A. R., & Lattimore, P. K. (2015). No-contact orders, victim safety, and offender recidivism in cases of misdemeanor criminal domestic violence: A randomized experiment. *American

Journal of Criminal Justice, 40(2), 225-249. https://doi.org/10.1007/s12103-014-9242-x

*Coulter, M., & VandeWeerd, C. (2009). Reducing domestic violence and other criminal recidivism: Effectiveness of a multilevel batterers intervention program. *Violence and Victims, 24*(2), 139-152. https://doi.org/10.1891/0886-6708.24.2.139

*Davis, R. C., Taylor, B. G., & Maxwell, C. D. (1998). Does batterer treatment reduce violence? A randomized experiment in Brooklyn. *Justice Quarterly, 18*(1), 171-201. https://doi.org/10.1080/07418820100094861

*Dobash, R., Dobash, R. E., Cavanagh, K., & Lewis, R. (1996). *Reeducation programs for violent men: An evaluation* (Research Findings, No. 46). http://webarchive.nationalarchives.gov.uk/20110314171826/http://rds.homeoffice.gov.uk/rds/pdfs/r46.pdf

*Dunford, F. W. (2000). The San Diego Navy experiment: An assessment of interventions for men who assault their wives. *Journal of Consulting and Clinical Psychology, 68*(3), 468-476. https://doi.org/10.1037/0022-006x.68.3.468

European Union Agency for Fundamental Rights. (2014). *Violence against women: An EU-wide survey. Results at a glance.* https://fra.europa.eu/sites/default/files/fra-2014-vaw-survey-at-a-glance-oct14_en.pdf

Fariña, F., Redondo, L., Seijo, D., Novo, M., & Arce, R. (2017). A meta-analytic review of the MMPI validity scales and indexes to detect defensiveness in custody evaluations. *International Journal of Clinical and Health Psychology, 17*, 128-138. https://doi.org/10.1016/j.ijchp.2017.02.002

*Feder, L. R., & Dugan, L. (2004). *Testing a court-mandated treatment program for domestic violence offenders: The Broward experiment.* National Institute of Justice. Washington, DC. https://www.ncjrs.gov/pdffiles1/nij/199729.pdf

Feder, L., & Wilson, D. B. (2005). A meta-analytic review of court-mandated batterer intervention programs: Can courts affect abusers' behavior? *Journal of Experimental Criminology, 1*(2), 239-262. https://doi.org/10.1007/s11292-005-1179-0

Gallego, R., Novo, M., Fariña, F., & Arce, R. (2019). Child-to-parent violence and parent-to-child-violence: A meta-analytic review. *European Journal of Psychology Applied to legal Context.* Advance online publication. https://doi.org/10.5093/ejpalc2019a4

Gondolf, E. W. (2000). Reassault at 30-months after batterer program intake. *International Journal of Offender Therapy and Comparative Criminology, 44*(1), 111-128. https://doi.org/10.1177/0306624x00441010

*Graña, J. L., Redondo, N., Muñoz-Rivas, M. J., & Cuenca, M. L. (2017). Short-term treatment response, attrition and recidivism in a partner violent men typology compared with a control group. *Psicología Conductual, 25*(3), 465-482.

*Haggård, U., Freij, I., Danielsson, M., Wenander, D., & Långström, N. (2017). Effectiveness of the IDAP treatment program for male perpetrators of intimate partner violence: A controlled study of criminal recidivism. *Journal of Interpersonal Violence, 32*(7), 1027-1043. https://doi.org/10.1177/0886260515586377

*Hasisi, B., Shoham, E., Weisburd, D., Haviv, N., & Zelig, A. (2016). The "care package" prison domestic violence programs and recidivism: A quasi-experimental study. *Journal of Experimental Criminology, 12*(4), 563-586. https://doi.org/10.1007/s11292-016-9266-y

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Academic Press.

*Herman, K., Rotunda, R., Williamson, G., & Vodanovich, S. (2014). Outcomes from a Duluth model batterer intervention program at completion and long term follow-up. *Journal of Offender Rehabilitation, 53*(1), 1-18. https://doi.org/10.1080/10509674.2013.861316

Holt, V. L., Kernic, M. A., Wolf, M. E., & Rivara, F. P. (2003). Do protection orders affect the likelihood of future partner violence and injury? *American Journal of Preventive Medicine, 24*, 16-21. https://doi.org/10.1016/S0749-3797(02)00576-7

*Jenkins, J. A., & Menton, C. (2003). The relationship between incarcerated batterers' cognitive characteristics and the effectiveness of behavioral treatment. *Corrections Compendium, 28*(3), 1-27. http://wheelsforjustice.rwu.edu/publications/CorrectionsCompendium-Mar2003.pdf

*Jones, A. S., & Gondolf, E. W. (2002). Assessing the effect of batterer program completion on reassault: An instrumental variables analysis. *Journal of Quantitative Criminology, 18*(1), 71-98. https://doi.org/10.1023/a:1013244929733

*Labriola, M., Rempel, M., & Davis, R. C. (2005). *Testing the effectiveness of batterer programs and judicial monitoring. Results from a randomized trial at the Bronx Misdemeanor Domestic Violence Court.* Center for Court Innovation. Nueva York, NY. http://www.courtinnovation.org/sites/default/files/battererprogramseffectiveness.pdf

Levesque, D. A., & Gelles, R. J. (1998, July). *Does treatment reduce recidivism in men who batter? A meta-analytic evaluation of treatment outcome.* Paper presented at the Program Evaluation and Family Violence Research: An International Conference, Durham, NH.

Ley Orgánica 1/1979, de 26 de septiembre, General Penitenciaria. (1979). *Boletín Oficial del Estado, 281,* 33987-34058. https://www.boe.es/buscar/pdf/1995/BOE-A-1995-25444-consolidado.pdf

*Lila, M., Gracia, E., & Catalá-Miñana, A. (2018). Individualized motivational plans in batterer intervention programs: A randomized clinical trial. *Journal of Consulting and Clinical Psychology, 86*(4), 309-320. https://doi.org/10.1037/ccp0000291

Lila, M., Martín-Fernández, M., Gracia, E., López-Ossorio, J. J., & González, J. L. (2019). Identifying key predictors of recidivism among offenders attending a batterer intervention program: A survival analysis. *Psychosocial Intervention, 28*(3), 157-167. https://doi.org/10.5093/pi2019a19

*Lin, S., Su, C., Chou, F. H., Chen, S., Huang, J., Wu, G. T., Chen, W. J., Chao, S. S., & Chen, C. (2009). Domestic violence recidivism in high-risk Taiwanese offenders after the completion of violence treatment programs. *Journal of Forensic Psychiatry & Psychology, 20*(3), 458-472. https://doi.org/10.1080/14789940802638341

Martín, A. M., Padrón, F., & Redondo, S. (2019). Early narratives of desistance from crime in different prison regimes. *European Journal of Psychology Applied to Legal Context, 11*(2), 71-79. https://doi.org/10.5093/ejpalc2019a2

Martín, B., & Moral, M. V. (2019). Relación entre dependencia emocional y maltrato psicológico en forma de victimización y agresión en jóvenes [Relationship between emotional dependence and psychological abuse in the form of victimization and aggression in young people]. *Revista Iberoamericana de Psicología y Salud, 10*(2), 75-89. https://doi.org/10.23923/j.rips.2019.02.027

Martín-Fernández, M., Gracia, E., & Lila, M. (2018). Assessing victim-blaming attitudes in cases of intimate partner violence against women: Development and validation of the VB-IPVAW scale. *Psychosocial Intervention, 27*(3), 133-143. https://doi.org/10.5093/pi2018a18

Martín-Fernández, M., Gracia, E., Marco, M., Vargas, V., Santirso, F. A., & Lila, M. (2018). Measuring acceptability of intimate partner violence against women: Development and validation of the A-IPVAW scale. *European Journal of Psychology Applied to Legal Context, 10*(1), 26-34. https://doi.org/10.5093/ejpalc2018a3

Maruna, S. (2004). Desistance and explanatory style: A new direction in the psychology of reform. *Journal of Contemporary Criminal Justice, 20,* 184-200. https://doi.org/10.1177/1043986204263778

Monteiro, A., Vázquez, M. J., Seijo, D., & Arce, R. (2018). ¿Son los criterios de realidad válidos para clasificar y discernir entre memorias de hechos auto-experimentados y de eventos vistos en vídeo? [Are the reality criteria valid to classify and to discriminate between memories of self-experienced events and memories of video-observed events?]. *Revista Iberoamericana de Psicología y Salud, 9*(2), 149-160. https://doi.org/10.23923/j.rips.2018.02.020

Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika, 8*(3), 161-168. https://doi.org/10.1007/BF02288700

*Murphy, C. M., Musser, P. H., & Maton, K. I. (1998). Coordinated community intervention for domestic abusers: Intervention system involvement and criminal recidivism. *Journal of Family Violence, 13*(3), 263-284. https://doi.org/10.1023/a:1022841022524

Novo, M., Fariña, F., Seijo, D., & Arce, R. (2012). Assessment of a community rehabilitation programme in convicted male intimate-partner violent offenders. *International Journal of Clinical and Health Psychology, 12*(2), 219-234. http://www.aepc.es/ijchp/articulos_pdf/ijchp-408.pdf

*Pérez, M. Giménez-Salinas A., & Juan, M. (2012). *Evaluación del programa "Violencia de género: programa de intervención para agresores", en medidas alternativas.* Instituto de Ciencias Forenses y de la Seguridad y Secretaría General de Instituciones Penitenciarias. Madrid, Spain. http://www.institucionpenitenciaria.es/web/export/sites/default/datos/descargables/descargas/VDG_EVALUACION_AUTONOMA_NIPO.pdf

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method bias in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879

Redondo, S., Sánchez-Meca, J., & Garrido, V. (2001). Treatment of offenders and recidivism: Assessment of the effectiveness of programmes applied in Europe. *Psychology in Spain, 5*(1), 47-62. http://www.psychologyinspain.com/content/reprints/2001/6.pdf

Redondo, S., Sánchez-Meca, J., & Garrido, V. (2002). Los programas psicológicos con delincuentes y su efectividad: la situación europea [Psychological programmes with offenders and their effectiveness: The European situation]. *Psicothema, 14*(Supl.), 164-173. http://www.psicothema.com/pdf/3487.pdf

Rosenfeld, B. D. (1992). Court ordered treatment of spouse abuse. *Clinical Psychology Review, 12*(2), 205-226. https://doi.org/10.1016/0272-7358(92)90115-o

Rosnow, R. L., & Rosenthal, R. (1988). Focused tests of significance and effect size estimation in counseling psychology. *Journal of Counseling Psychology, 35,* 203-208. https://doi.org/10.1037/0022-0167.35.2.203

*Saunders, D. G. (1996). Feminist-cognitive-behavioral and process-psychodynamic treatments for men who batter: Interaction of abuser traits and treatment models. *Violence and Victims, 11*(4), 393-414. https://doi.org/10.1891/0886-6708.11.4.393

Schmidt F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist, 36*(10), 1128-1137. https://doi.org/10.1037/0003-066X.36.10.1128

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting errors and bias in research findings* (3th ed.). Sage.

Schmucker, M., & Lösel, F. (2015). The effects of sexual offender treatment on recidivism: An international meta-analysis of sound quality

evaluations. *Journal of Experimental Criminology, 11*(4), 597-630. https://doi.org/10.1007/s11292-015-9241-z

*Stewart, L. A., Gabora, N., Kropp, P. R., & Lee, Z. (2014). Effectiveness of risk-needs-responsivity-based family violence programs with male offenders. *Journal of Family Violence, 29*(2), 151-164. https://doi.org/10.1007/s10896-013-9575-0

*Stith, S. M., Rosen, K. H., & McCollum, E. E. (2004). Treating intimate partner violence within intact couple relationships: Outcomes of multi-couple versus individual couple therapy. *Journal of Marital and Family Therapy, 30*(3), 305-318. https://doi.org/10.1111/j.1752-0606.2004.tb01242.x

Stover, C. S. (2005). Domestic violence research what have we learned and where do we go from here? *Journal of Interpersonal Violence, 20*(4), 448-454. https://doi.org/10.1177/0886260504267755

Taft, C. T., Murphy, C. M., Elliott, J. D., & Morrel, T. M. (2001). Attendance-enhancing procedures in group counseling for domestic abusers. *Journal of Counseling Psychology, 48*(1), 51-60. https://doi.org/10.1037/0022-0167.48.1.51

*Taylor, B. G., & Maxwell, C. D. (2009). *The effects of a short-term batterer treatment program for detained arrestees: A randomized experiment in the Sacramento County, California jail.* Department of Justice, Crime and Violence Prevention Center. Sacramento, CA. https://www.ncjrs.gov/pdffiles1/nij/grants/228275.pdf

*Tollefson, D. R., & Gross, E. (2006). Predicting recidivism following participation in a batterer treatment program. *Journal of Social Service Research, 32*(4), 39-62. https://doi.org/10.1300/j079v32n04_03

*Tollefson, D. R., Webb, K., Shumway, D., Block, S. H., & Nakamura, Y. (2009). A mind-body approach to domestic violence perpetrator treatment: Program overview and preliminary outcomes. *Journal of Aggression, Maltreatment & Trauma, 18*(1), 17-45. https://doi.org/10.1080/10926770802610657

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In I. Olkin, J. G. Ghurye, W. Hoeffding, W. G. Madoo, & H. Mann (Eds.), *Contributions to probability and statistics* (pp. 448-485). Stanford University Press.

Vilariño, M., Amado, B. G., Vázquez, M. J., & Arce, R. (2018). Psychological harm in women victims of intimate partner violence: Epidemiology and quantification of injury in mental health markers. *Psychosocial Intervention, 27*, 145-152. https://doi.org/10.5093/pi2018a23

Vilariño, M., Fariña, F., & Arce, R. (2009). Discriminating real victims from feigners of psychological injury in gender violence: Validating a protocol for forensic settings. *European Journal of Psychology Applied to Legal Context, 1*(2), 221-243. http://sepjf.webs.uvigo.es/index.php?option=com_docman&task=doc_download&gid=19&Itemid=110&lang=en

Appendix

Primary Studies

Author(s) and publication year	n	Treatment type and intervention format	Duration and length of intervention	Measure of recidivism during follow-up	Recidivism rate %		Efect size: δ	
Saunders (1996)	EG_1: 61 EG_2: 68	EG1: Duluth+ CBTP EG2: Psychodinamical process group	12 sessions + 20 support (32 weeks)	24 months (OR-CR)	OR EG_1: 23.2% EG_2: 20.3%	CR EG_1: 34% EG_2: 33.3%	OR -0.04 0.06	CR -0.17 -0.15
Dobash et al. (1996)	40	Duluth Group	6-7 months	12 months (OR-CR)	OR: 7%	CR: 33%	OR: 0.7	CR: -0.14
Murphy et al. (1998)	235	Duluth (format non specified)	22 sesssions	12-18 months (OR)	OR: 15.7%		0.23	
Babcock & Steiner (1999)	106	Duluth Group	36 weeks	24 months (OR)	OR: 8%		0.63	
Davis et al. (2000)	EG_1: 129 EG_2:61 CG: 186	Duluth Group	39 hours EG_1: 8 weeks EG_2: 26 weeks CG: TBC	6 and 12 months (OR-CR)	OR 6 m EG_1: 7% GG_2: 15% CG: 22% CR 6 m EG_1: 23% EG_2: 19% CG: 21%	OR 12 m EG_1: 10% EG_2: 25% CG: 26% CR 12 m EG_1: 14% EG_2: 18% CG: 22%	OR 6 m EG_1: 0.7 EG_2: 0.26 CR 6 m EG_1: 0.16 EG2: 0.29	OR 12 m EG_1: 0.51 EG_2: -0.09 CR 12 m EG_1: 0.49 EG2: 0.33
Dunford (2000)	EG_1: 168 EG_2: 153 EG_3: 173 CG: 150	CBTP EG_1 and EG_2: Group EG_3: Individual	EG_1: 12 months (6 weekks + 6 months) EG_2: 26 weeks + 6 months EG_3: 12 months	6 months (CR) 12 months (OR)	OR EG_1: 4% EG_2: 3% EG_3: 6% CG: 4%	CR EG_1:29 % EG_2: 30% EG_3: 27% CG: 35%	OR EG_1: 0.98 EG_2: 1.11 EG_3: 0.78	CR EG_1: -0.03 EG_2: -0.06 EG_3: 0.03
Jones and Gondolf (2002)	EG_{P1}: 213 EG_{P2}: 208 EG_{P3}: 215 EG_{P4}: 217	Duluth Group	EG_{P1}: 3 months EG_{P2}: 3 months EG_{P3}: 5.5 months EG_{P4}: 9 months	15 y 30 months (OR)	RO EG_{P1} EG_{P2} EG_{P3} EG_{P4}	15m 31.5% 31.8% 26.2% 24.7%	30m 41.2% 38.6% 34.2% 28.2	RO EG_{P1} EG_{P2} EG_{P3} EG_{P4} 15m -0.29 -0.3 -0.13 -0.09 30 m -0.55 -0.48 -0.36 -0.19
Jenkins and Menton (2003)	EG: 114	CBTP Group	9 weeks	30 months (OR)	OR: 10%		0.51	
Feder and Dugan (2004)	EG: 404 CG: 229	EG: Duluth Model	26 weeks	12 months (OR)	EG: 24% CG: 21%		-0.07	
Stith et al. (2004)	EG_1: 14 EG_2: 16 CG: 9	Duluth+ CBTP EG_1 Couple individual EG_2: Couple group	6 weeks	6 & 24 months (CR)	6 months EG_1: 43% EG_2: 25% CG: 67%	24 months EG_1: 0% EG_2: 13% CG: 50%	6 months EG_1: -0.41 EG_2: 0.09	24 months EG_1: 1.74 EG_2: 0.54
Bowen et al. (2005)	EG: 86	Duluth Group	24 weeks + 5 sessions	11 months (OR)	OR: 20%		0.03	
Bennet et al. (2005)	EG: 384	Duluth Group	24 semanas	18 months (OR)	OR: 15.4%		0.25	
Labriola et al. (2005)	EG: 157	Duluth Group	26 weeks	12 months (OR)	OR: 6%		0.78	
Tolleffson and Gross (2006)	EG: 102	Duluth Group	20 sessions	7-58 months (OR)	OR: 18%		0.14	
Tollefson et al. (2008)	EG: 57	Mind-body Bridging Group	8-10 sessions	9-27 months (OR)	OR: 9%		0.57	
Lin et al. (2009)	EG: 70 CG: 231	Duluth+ CBTP Group	12-18 weeks	6 y 9 months (CR)	6 months EG:34.3% CG:34.2%	9 months EG:27.1% CG:21.2%	6 months 0.18	9 months 0.03
Taylor and Maxwell (2009)	EG:317 CG: 312	Duluth Group	5 days	6 and 12 months (OR)	6 months EG:65.9% CG:65.7%	12 months EG:68.6% CG:69.6%	6 months -1.18	12 months -1.26
Coulter and VandeWeerd (2009)	EG_1:1424 EG_2: 9386 EG_3:1712	G_1 and G_2: Duluth Group G_3: Specialized treatment	EG_1: 8-12 weeks EG_2: 26 weeks EG_3: 26 weeks-1year	1-10 years (OR)	OR EG_1: 8.8% EG_2: 8.3% EG_3: 8.6%		OR EG_1: 0.58 EG_2: 0.61 EG_3: 0.59	
Pérez et al. (2012)	EG: 598	CBTP Group	25 weeks	12 months (OR)	OR: 4.6%		0.91	
Herman et al. (2014)	EG: 103 CG: 53	Duluth (format not specified)	24 weeks	9 years (OR)	EG: 39% CG: 37.7%		-0.49	
Stewart et al. (2014)	EG_{HI}: 84 EG_{MI}: 76	RNR Group	EG_{HI}: 78 sessions EG_{MI}: 25 sessions	1 year aprox (OR)	EG_{HI}: 4% CG: 14%	EG_{MI}: 4% CG: 12%	EG_{HI}: 0.98 EG_{MI}: 0.98	

Appendix (continued)

Author(s) and publication year	n	Treatment type and intervention format	Duration and length of intervention	Measure of recidivism during follow-up	Recidivism rate %		Efect size: δ
Haggård et al. (2015)	EG: 37 CG: 28	Duluth Group	27 sessions	4 years aprox. (OR)	EG: 15% CG: 31%		0.26
Hasisi et al. (2016)	EG₁:247 CG:240 EG₂:204 CG: 195 EG₃:147 CG: 143 EG₄:101 CG: 98 EG₅: 53 CG: 63	The House of Hope (Therapeutic, group & individual)	Undefined	1 year (OR) 2 years (OR) 3 years (OR) 4 years (OR) 5 years (OR)	G1: 3.2% G2: 5.9% G3: 8.8% G4: 9.9% G5: 12.2%	CG: 8.8% CG:14.9% CG: 19.6% CG: 18.4% CG: 25.4%	1.08 0.79 0.58 0.51 0.34
Graña et al. (2017)	EG$_{HL}$: 19 EG$_{ML}$: 74 EG$_{LL}$: 173 CG: 44	CBTP Group	23 weeks	12 months (OR)	EG$_{HL}$: 5.3% EG$_{ML}$: 5.4% EG$_{LL}$: 4% CG:6.8%		EG$_{HL}$: 0.84 EG$_{ML}$: 0.83 EG$_{LL}$: 0.98
Lila et al. (2018)	EG: 64	CBTP + IMP	35 sessions	6 months (OR)	OR: 6.25%		0.76

Note. EG = experimental group; CG = control group; CBTP = cognitive-behavioral treatment program; IMP = individualized motivational plan; OR = official records; CR = couple records.