**Heterogeneity in interpersonal violence outcome research: an investigation and discussion of clinical and research implications**

JC Hockenhull1, MG Cherry2, R Whittington3, RC Dickson1, M Leitner4, W Barr3, J McGuire2

1. Liverpool Reviews and Implementation Group, University of Liverpool, Whelan Building, Brownlow Hill, Liverpool, L69 3GB

2. Division of Clinical Psychology, University of Liverpool, Whelan Building, Brownlow Hill, Liverpool, L69 3GB

3. Health and Community Care Research Unit, University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool, L69 7ZA

4. ER & IC, Medical Division, Macclesfield, Cheshire, UK

Correspondence to:

Juliet C. Hockenhull, Liverpool Reviews and Implementation Group, Room 2.07, Whelan Building, The University of Liverpool, The Quadrangle, Brownlow Hill, Liverpool, L69 3GE. **E** [**julieth@liverpool.ac.uk**](mailto:julieth@liverpool.ac.uk) **T 0151 7955441**

Paper based on a presentation at the 7th European Congress on Violence in Psychiatry, Prague, Czech Republic, October 2011

## Abstract

The term ‘interpersonal violence’ is used to refer to a wide range of violent acts, including intimate partner violence, child abuse and stranger assaults. Two extensive reviews of interventions to reduce or prevent interpersonal violence published between 1950 and 2008 identified a large literature base reflecting the extensive collective effort of violence researchers over the past 50 years. However, neither review was able to meaningfully synthesize sub-group data due to the high degree of heterogeneity present. This paper interrogates this apparent contradiction by examining three examples of predefined sub-group analyses from these reviews. None of the chosen examples produced groups of studies with adequate homogeneity for meaningful meta-analysis and synthesis, indicating that the violence research literature, while extensive, is currently too heterogeneous to be used to inform policy related to the most appropriate interventions suitable for evidence-based practice. If the literature cannot become more focused via a major topic prioritization exercise, an alternative solution to this problem may be to adopt a realist synthesis approach to determine what works, for whom, and in what context.

**Keywords**: systematic review; intervention; violence; homogeneity; meta-analysis

# Background

Violence is a global issue and raises ongoing public, political, and professional concern due to its significant impact on health and wellbeing in all societies. The World Health Organization’s (WHO) ([Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002](#_ENREF_13)) definition of violence includes threats, intimidation and physical, sexual, and psychological abuse, as well as acts of self-harm and suicidal behavior. Interpersonal violence (i.e., intimate partner violence, child abuse, and stranger assaults) is thought to account for more than 486,000 deaths per year worldwide ([World Health Organisation, 2013](#_ENREF_24)), and these are only a fraction of all assaults that occur.

Due to the ongoing concern over violence and its effects on individuals and society, the reduction and prevention of violence has been extensively researched over the past 50 years, which has led to the development and implementation of a broad range of interventions. This research has done much to improve our understanding of violence. However, it has also led to a diverse literature, reflective of the many different manifestations of violence, the different populations in which violence occurs, and the methods adopted to evaluate the efficacy of interventions to reduce violence or associated outcomes.

Such a diverse and extensive literature can be bewildering and can obscure the effectiveness of particular interventions to decrease violent behavior in specific situations. When practitioners are choosing which intervention to employ it is important that they consider using an evidence-based practice approach. However, such an evidence base can only be confidently constructed through consistent replication of positive findings of particular interventions with specific populations. This replication and consolidation is the hallmark of science and, without it, it is difficult to draw meaningful conclusions about the generalizability of findings beyond the particular study being reported.

Evidence-based practice is seen as a way of providing health care and other public services that is guided by a thoughtful integration of the best available scientific knowledge with clinical expertise. This approach allows the practitioner to critically assess research data, clinical guidelines, and other information resources in order to correctly identify the clinical problem, apply the most high-quality intervention, and re-evaluate the outcome for future improvement ([U.S. National Library of Medicine, 2009](#_ENREF_23)). However, given an extensive and often complex literature, practitioners rarely have the resources to collate all evidence but rather turn to reviews of research studies for guidance. Systematic review with, where possible, aggregation of data from a range of comparable studies through meta-analysis, is accepted as the approach necessary to establish what constitutes ‘the best available scientific knowledge’ on effectiveness at any particular time.

The synthesis of data through meta-analysis increases the strength of evidence and allows for researchers to test whether the findings of individual studies are robust. This is particularly true where either the numbers of participants in individual studies is small, and/or where a smaller intervention effect size is expected and may require large samples to be detectable. For data synthesis to be meaningful, studies need to be as homogenous (composed of similar or identical parts or elements) as possible otherwise any significant effects identified may be due to factors other than the treatment being tested. The level of homogeneity/heterogeneity between studies can be quantified using either the Q statistic (which assesses presence of heterogeneity) or I2 statistic (which measures the extent of heterogeneity).

It must be noted that systematic review and meta-analysis, while considered the most robust way to synthesize data, is not the only recognized method to synthesize the findings of individual intervention studies in a way which enables meaning to be drawn. For example, realist synthesis aims to review the evidence for complex interventions and identifies the context and mechanisms by which a particular outcome can be achieved. Rather than adopt a rigid and systematic approach to identifying all literature relating to a certain topic or intervention, researchers conducting realist synthesis adopt a more iterative approach. They aim to explore *how* complex programmes or interventions work, or *why* they fail, in certain circumstances or with certain population groups. In contrast to RCTs, whose purpose is to test for cause-effect relationships, realist synthesis adopts the premise that in order to infer a causal outcome between two events or processes (such as a CBT-based intervention to reduce violence, and subsequent reoffending rates), the mechanism that connects them and the context in which this mechanism occurs must also be understood. However, in the field of violence research, realist synthesis has largely been overlooked, with researchers tending to synthesize studies using more ‘traditional’ approaches. Indeed, a large number of systematic reviews and meta-analyses have been conducted over the past decade to summarize and integrate the findings from the literature. However, in line with the traditional requirement of a systematic review to include comparable studies, these reviews tend to focus on a specific intervention (e.g., second generation antipsychotics ([Bhana, Foster, Olney, & Plosker, 2001](#_ENREF_1)) and/or a specific outcome (e.g., re-offending ([Schmucker & Losel, 2008](#_ENREF_19))) in various singular populations (e.g., people with learning disability ([Hassiotis & Hall, 2004](#_ENREF_9))).

It is worth noting that many of these previous focused reviews were unable to meaningfully synthesize the results of relevant randomized controlled trials (RCTs; considered ‘gold standard’ evidence) due to high levels of heterogeneity between them ([Brooks-Gordon, Bilby, & Wells, 2006](#_ENREF_3); [Duncan, Nicol, Ager, & Dalgleish, 2006](#_ENREF_6); [Hassiotis & Hall, 2004](#_ENREF_9); [Huf, Alexander, & Allen, 2004](#_ENREF_11); [Smedslund, Dalsbo, Steiro, Winsvold, & Clench-Aas, 2007](#_ENREF_21)). For example, Schmucker and Lösel ([2008](#_ENREF_19)), in their systematic review of controlled outcome evaluations of psychosocial and organic sexual offender treatments, found “*too few randomized trials on sexual offender treatment with too heterogeneous modes of treatment as to carry out a differentiated analysis*” ([Schmucker & Losel, 2008](#_ENREF_19)) p16. In order to meaningfully synthesize study results, Schmucker and Lösel ([2008](#_ENREF_19)) had to include data from both randomized and non-randomized studies and adjust for heterogeneity by using moderator analyses, thus indicating the difficulties in meaningfully synthesizing data in the field of violence research.

A recent review conducted by the Liverpool Violence Group (LiVio) ([Leitner, Barr, McGuire, Jones, & Whittington, 2006](#_ENREF_14)) adopted a more comprehensive approach by applying systematic review methods to synthesize the results of research on all interventions published up to 2004 and relating to a broad range of violence-related outcomes amongst a wide mental health and criminal justice population. While this contravened the traditional focused review approach noted above and had significant resource implications, it was felt to be desirable as a way of capturing the full range of relevant work and helping to heal the split between the parallel clinical and criminological literatures. Hollin (2008) and others have argued that these literatures are fragmented and should be reintegrated where possible to the mutual benefit of practitioners and researchers in both settings. The review included 410 studies, of which 301 presented statistical analyses and were the main focus for data analysis. However, when data were limited to the best standard of evidence (RCTs), meta-analyses were not possible due to the level of heterogeneity in the reporting of findings, methodological approaches, interventions, comparators, populations, and settings. Based on the findings of the review, recommendations were made to researchers to encourage future work to be planned in such a way as to increase the degree of conceptual and methodological rigor and the potential for replication ([Leitner et al., 2006](#_ENREF_14)).

A subsequent update of this review, covering publications from 2002-2008, has now been completed ([Hockenhull et al., 2012](#_ENREF_10); [Leitner et al., 2006](#_ENREF_14)), using the same search strategy and databases as Leitner et al. (2006) wherever possible. The update identified 195 studies of interventions meeting the inclusion criteria, of which 40 were RCTs. As with the original review, these studies covered a wide range of populations, interventions and designs. An initial meta-analysis of all 40 identified RCTs showed a significant treatment odds ratio effect of 0.35 (95% CI 0.26, 0.49) (random effect model in which the odds ratio (OR) <1 favors treatment over control) (Hockenhull et al., 2012). Many researchers would view this finding as relatively robust. However, the heterogeneity in the sample (I2 = 86%) was extremely large, and subsequent meta-analyses of RCT sub-groups (i.e., by population, comparator,etc.) demonstrated that this variance was not a function simply of either intervention or outcome type, with most sub-group analyses still showing high levels of heterogeneity. Thus there was insufficient consistency within the effect sizes to permit any conclusions to be drawn about which methods could be recommended as more reliably producing benefits.

The two LiVio reviews ([Hockenhull et al., 2012](#_ENREF_10); [Leitner et al., 2006](#_ENREF_14)), when combined together, represent a comprehensive database containing evidence published between 1950 and 2008 relating to interventions designed to influence a broad range of violence-related outcomes in a wide mental health and criminal justice population. However, in line with previous reviews, the findings of the two LiVio group reviews ([Hockenhull et al., 2012](#_ENREF_10); [Leitner et al., 2006](#_ENREF_14)) indicate large levels of heterogeneity across studies, which cannot be adequately explained by differences in either the type of intervention or outcome measure used. The sub-group analyses performed in Hockenhull et al. ([2012](#_ENREF_10)) considered one of these characteristics at a time, for example by analyzing data as a function of type of intervention. It is possible that heterogeneity can be reduced and more robust findings established by amalgamating the evidence from the LiVio group reviews and identifying reasonably-sized groups of comparable studies with adequate homogeneity and statistical power.

Having become familiar with this literature, the LiVio group was concerned to discover that despite its quantity and range of coverage, it contained surprisingly few sets of studies with sufficient similarity to each other to permit firm conclusions to be drawn regarding the impact of interventions. The following objective for the present study was, therefore, set: to test the literature with respect to the extent of homogeneity or heterogeneity that could be found when seeking to answer selected domain-limited questions of practical interest. The research question explored in this paper was, therefore: When stringent criteria are applied to the violence research literature, can studies be meaningfully compared and aggregated in a way that allows practical conclusions to be drawn? This question was addressed by applying stringent criteria adapted from a systematic review framework, relating to population, intervention, comparator, and outcome (PICO).

# Methods

Systematic review methodology was used to conduct the two LiVio systematic reviews, full details of which can be found in the full-text report of each review ([Hockenhull et al., 2012](#_ENREF_10); [Leitner et al., 2006](#_ENREF_14)). For each review, relevant data were extracted from each included paper and entered into an SPSS database. Extracted data included, but were not limited to data pertaining to population, setting, type of offense, intervention received, comparator(s), duration of intervention, outcome measure(s) and numerical outcome data ([Hockenhull et al., 2012](#_ENREF_10); [Leitner et al., 2006](#_ENREF_14)).

For purposes of this paper, the two SPSS data extraction datasheets from the two reviews were merged into one SPSS file and exported into Microsoft Access. Due to the diversity in included studies, and, therefore, the immense number of subgroup analyses possible, three examples of stringent PICO criteria were chosen on an ad hoc basis by two authors (JCH and RW) to address the research question. Specific inclusion criteria covering population, intervention, comparator, outcome, and design aspects for the three different examples of relevant clinical and/or offender sub-group analyses conducted are displayed in Table 1.

*Table 1: PICO values for each of the three examples*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Example 1** | **Example 2** | **Example 3** |
| Population | Individuals with a diagnosis of schizophrenia or schizo-affective disorder | Sex offenders | Domestic violence perpetrators |
| Intervention | Atypical antipsychotics | CBT | CBT |
| Comparison | Other atypical antipsychotics | TAU | No treatment |
| Outcome | Scale measures | Reconviction or reoffending rates | Reconviction or reoffending rates |
| Study design | RCTs | Quasi-experimental | RCTs |

* RCT= randomised controlled trial; TAU = treatment as usual; CBT= cognitive behavioural therapy

For each example, the relevant variables in the Microsoft Access database were filtered in a stepwise manner and the number of studies meeting the inclusion criteria at each step was recorded. Once all filters had been applied, details of included studies were then examined for each example, and the appropriateness of meta-analysis considered. Where appropriate, meta-analyses were conducted and the heterogeneity was assessed using the I2 statistic.

# Results

A total of 496 studies of interventions to reduce or prevent violence were included in the database (301 from the original review plus 195 from the update). The references for these studies are available from the authors upon request. Results for each example will now be separately discussed.

## Example 1: Randomized controlled trials evaluating atypical antipsychotics to reduce violence in individuals on the schizophrenic spectrum (outcome measured using scale measures)

The PICO criteria for Example 1 were successively applied to the 496 studies in the database. Sixty-one studies focused on patients with schizophrenia or schizo-affective disorder (POPULATION); 53 of these studied a pharmacological intervention, of which 25 evaluated the use of atypical antipsychotics (INTERVENTION). The other interventions being studied and therefore excluded from this sub-group included traditional antipsychotics (n = 28), anti-epileptic drugs (n = 5), “other” types of drugs (n = 4), beta blockers (n = 3), anti-manic drugs (n = 3), and antidepressants (n = 1).

Of the 25 studies evaluating the use of atypical antipsychotics, eight compared two or more active treatments (COMPARATOR). Other comparators included a placebo (n = 5), different doses of an active treatment (n = 2), and treatment as usual (TAU) (n = 1). The nature of this comparison is important methodologically as it is easier to test whether or not there is an effect against a placebo or no treatment than it is against TAU or some form of active treatment; it is, therefore, important that comparisons are made consistently. A further 11 of the 25 studies did not use any comparator group but were uncontrolled pre/post test studies and thus were excluded from analysis.

Six of the eight studies comparing atypical antipsychotics to another active treatment used a scale outcome measure and thus fulfilled the defined criteria for outcome (OUTCOME). Of these six studies, three were excluded based on design (two were quasi-experimental and one used a crossover design) (STUDY DESIGN). Therefore, three out of the original 496 studies (0.6%) met the PICO inclusion criteria outlined in Table 1. Figure 1 (below) illustrates the stepwise selection process when the PICO criteria were applied.

*Figure 1: Number of studies included at each stage of application of the PICO table variables (Example 1)*

The three apparently homogenous selected studies were then examined in more detail (see Table 2).

*Table 2: Details of included studies (Example 1)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Author** | | |
|  | | **Citrome et al. (2001)** | **Czobor et al. (1995)** | **Kane et al. (2001)** |
| Participant characteristics | | Patients meeting DSM-IV criteria for schizophrenia or schizo-affective disorder  History of sub-optimal treatment response  Aged between 18 and 60 years | Patients meeting DSM-III-R criteria for schizophrenia  Aged between 18 and 65 years | Patients meeting DSM-III-R criteria for schizophrenia or schizo-affective disorder  Partially responsive patients enrolled  Aged between 20 and 55 years |
| Total N (all groups) | | 157 | 139 | 71 |
| Study setting | | Open in-patient hospital ward | Open in-patient hospital ward | Mixed settings |
| Drug evaluated | Clozapine | ✓ | x | ✓ |
|  | Olanzapine | ✓ | x | X |
|  | Risperidone | ✓ | ✓ | X |
|  | Haloperidol | ✓ | ✓ | ✓ |
|  | Dose (mg) | Clozapine = 500 (range 200-800)  Olanzapine = 20 (range 10-40)  Risperidone = 8 (range 4-16)  Haloperidol = 20 (range 10-30) | Risperidone = 2, 6, 10 or 16  Haloperidol =20 | Clozapine = 500  Haloperidol =10 |
| Follow up (weeks) | | 1, 2, 3, 4, 6, 8, 10, 12, 14 | 2, 4, 6, 8 | 5, 11, 17, 29 |
| Outcome measure used | | PANSS hostility | PANSS hostility | BPRS hostility |
| Data type | | Mean scores | Mean change | Mean scores |

* PANSS hostility = hostility item of the Positive and Negative Syndrome Scale; BPRS hostility = hostility item of the Brief Psychiatric Ratings Scale

It is clear that the type of atypical antipsychotics being investigated differed between the studies, although two sets of studies used the same comparators, allowing potential comparisons to be made (haloperidol vs. risperidone studied by Citrome 2001 and Czobor 1995 and clozapine vs. haloperidol by Citrome 2001 and Kane 2001).

However, when other details of the studies were examined, it became clear they differed in terms of population, outcome measures, dosing schedules, follow up periods, and setting. Furthermore, the outcome metric in the research studies differed with either mean absolute baseline/endpoint scores or mean relative baseline/endpoint changes being reported. Whilst the latter mean can be calculated from the former, standard deviations for the recalculated change score cannot be calculated. Meta-analysis was, therefore, not considered appropriate in this instance.

## Example 2: Quasi-experimental studies evaluating cognitive behavioral therapy compared with treatment as usual to reduce violence in sex offenders (outcome measured using reconviction or reoffending rates)

The PICO criteria for Example 2 were then successively applied to the 496 studies in the database. Of these studies, individuals who had committed sexual offenses were the focus of 56 (POPULATION). Forty-six of these 56 studies evaluated the use of a psychological intervention, of which 24 were based on a CBT approach (INTERVENTION). The other approaches used in the excluded studies were multi-modal (n = 9), “other” (n = 7), psychotherapy (n = 3), behavior modification (n = 2), and skills training approach (n = 1). Fourteen of the 24 CBT based studies used some type of group comparison (although four of these compared different groups post-hoc, such as completers vs. non-completers). Seven compared the treatment against TAU, two compared two or more active treatments, and one a placebo. For the purposes of this example, studies comparing an intervention against TAU (n = 7) were selected (COMPARATOR).

All seven of these studies used reconviction or re-offending rates as their outcome (OUTCOME) and four studies employed a quasi-experimental design (STUDY DESIGN). These four studies ([Friendship, Mann, & Beech, 2003](#_ENREF_8); [McGrath, Hoke, & Vojtisek, 1998](#_ENREF_16); [Nicholaichuk, Gordon, Gu, & Wong, 2000](#_ENREF_17); [Schweitzer & Dwyer, 2003](#_ENREF_20)) were included in the analysis. The remaining three studies were linked papers of a RCT and were thus excluded. The number of studies included at each stage of the application of inclusion criteria are shown in Figure 2.

*Figure 2: Number of studies included at each stage of application of the PICO table variables (Example 2)*

Details of the four included studies are shown in Table 3.

*Table 3: Details of included studies (Example 2)*

|  | **Author** | | | |
| --- | --- | --- | --- | --- |
|  | **Friendship et al. (2003)** | **McGrath et al. (1998)** | **Nicholaichuk et al. (2000)** | **Schweitzer et al. (2003)** |
| Population characteristics | Adult male sex offenders serving > 4 years in prison for sexual offense  Volunteered for intervention | Adult male sex offenders  On community supervision for ≥ 3 months during study period | High risk adult male sex offenders  Volunteered for intervention  Released into community post-intervention | Convicted adult male sex offenders  Released into community post-intervention |
| Total N (all groups) | 2557 | 122 | 374 | 445 |
| Intervention details | SOTP  29 two-hour sessions, subsequently expanded to 70-75 sessions  Prisoners completed either one or other program | Specialized treatment program for 3 months  CBT based | CBT-based intervention (no other details provided) | CBT-based SOTP consisting of:  15-week assessment and treatment phase  5-week treatment planning phase  25-week intensive treatment phase |
| Comparator details | Prisoners matched to intervention group by year of discharge  No other details given | Comparator A: Non-specialized treatment group  Comparator B: No-treatment group (individuals who had refused treatment) | Stratified matched comparison group of incarcerated sexual offenders drawn from archives | Comparator A: Non-completing sex offenders  Comparator B: Sex offenders who received no treatment in prison |
| Outcome data type | Reconviction for sexual offence | Reconviction for sexual offence | Readmission to prison for sexual offence | Reconviction for sexual offence |
| Start setting | Penal institution | Community | Secure forensic mental health setting | Penal institution |
| End setting | Community | Community | Secure forensic mental health setting | Community |
| Average follow up (months) | 24 | Intervention: 52.9  Comparator A: 63.8  Comparator B: 65.2 | Intervention: 70.8  Comparator: 87.6 | Intervention: 58  Comparator A: 66  Comparator B: 62 |
| Comments | Comparison group were retrospectively selected  Baseline characteristics differed between groups  Combined two treatment programs both of which the authors point out that by today’s standards lacked sophistication | Significant differences in baseline characteristics between no-treatment group and other groups |  |  |

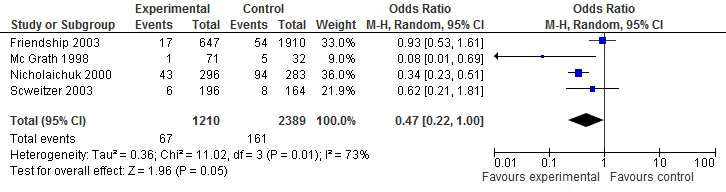
SOTP = Sex offender treatment program, CBT = cognitive behavioural therapy

Unlike Example 1, it was considered feasible to aggregate some studies and therefore the results of each study are summarized in Table 4.

*Table 4: Results of included studies: reconviction rates (Example 2)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Author** | | | |
| **Reconviction rates, n (%)** | **Friendship et al., 2003** | **McGrath et al., 1998** | **Nicholaichuk et al., 2000** | **Schweitzer et al., 2003** |
| Treatment group | 17/647 (2.6) | 1/71 (1.4) | 43/296 (14.5) | 6/196 (3.1) |
| Comparison group A | 54/1910 (2.8) | 5/32 (15.6) | 94/283 (33.2) | 6/85 (7.1) |
| Comparison group B | - | 2/19 (10.5) | - | 8/164 (4.9) |
| Significant reduction in reconviction rates when intervention group compared with comparator? | No | Yes | Yes | No |

Two studies ([McGrath et al., 1998](#_ENREF_16); [Nicholaichuk et al., 2000](#_ENREF_17)) found a significant reduction in reconviction rates when the intervention group was compared with comparator(s) and two studies did not ([Friendship et al., 2003](#_ENREF_8); [Schweitzer & Dwyer, 2003](#_ENREF_20)). The reconviction rates across studies varied considerably, with Friendship et al. (2003) reporting 2.8% in the comparison group compared to 33.2% reported by Nicholaichuk et al. (2000). In the intervention groups, reconviction rates ranged between 1.4% and 14.5% across studies. Despite these differences, conduct of meta-analysis was considered appropriate. Figure 3 shows the forest plot for the meta-analysis.



*Figure 3: Meta-analysis of included studies (Example 2)*

The overall test for effect reached statistical significance (*Z* = 1.96, *p* = 0.05) but heterogeneity (I2) was moderately high at 73%. Sensitivity analyses showed that the heterogeneity was reduced to 31% when the [Friendship et al., 2003](#_ENREF_8) study was removed (*Z* = 3.12, *p* = 0.002). This study differed from the others most notably on the length of follow up period; which at 24 months was shorter than 63.8 ([McGrath et al., 1998](#_ENREF_16)), 87.6 ([Nicholaichuk et al., 2000](#_ENREF_17" \o "Nicholaichuk, 2000 #21)), and 62 ([Schweitzer & Dwyer, 2003](#_ENREF_20)) months.

Therefore, despite the careful application of inclusion criteria, the meta-analysis still showed that the included studies were heterogeneous and therefore that meaningful synthesis and interpretation of the combined results of these studies was difficult.

## Example 3: Randomized controlled trials evaluating the use of cognitive behavioral therapy compared with no treatment to reduce violence in domestic violence perpetrators (outcome measured using criminal statistics)

Finally, the PICO criteria for Example 3 were successively applied to the 496 studies in the database. Sixty-nine studies investigated domestic violence perpetrators (POPULATION) and 55 of these applied a psychological intervention, of which 23 evaluated the use of CBT-based interventions (INTERVENTION). Other intervention approaches included behavior modification (n = 9), multiple models (n = 3), educational approach (n = 2) psychotherapy (n = 2) and skills training (n = 1). Two further studies did not state their approach and 13 used approaches other than those named here. Three of the studies evaluating CBT-based interventions compared them against no treatment (COMPARATOR). The other 20 excluded studies compared: different groups (primarily completers and non-completers of CBT-based interventions) (n=6), two or more active treatments (n = 5), and CBT against TAU (n = 1). The remaining eight studies used a pre/post design and, therefore, did not have a comparator group.

Of the three remaining studies ([Bresnahan, 1999](#_ENREF_2); [Dunford, 2000](#_ENREF_7); [Ley, 2006](#_ENREF_15)), two used reconviction or reoffending rates as an outcome and two studies a scale measure (OUTCOME). The study using both was an RCT that also used self-report and reports by significant others ([Dunford, 2000](#_ENREF_7)), whereas the other two were quasi-experimental designs ([Bresnahan, 1999](#_ENREF_2); [Ley, 2006](#_ENREF_15)) (DESIGN). Therefore, any selection on both outcome and design would lead to only one study being included and thus no meta-analysis would be possible or necessary. Figure 4 shows the number of studies selected at each stage of the application of the PICO inclusion criteria, resulting in only one final included study. For illustration, details of the three studies investigating the efficacy of CBT-based interventions compared with no treatment in a domestic violence population are shown in Table 5.

*Figure 4: Number of studies included at each stage of application of the PICO table variables (Example 3)*

*Table 5: Details of three studies comparing a cognitive behavioural therapy based intervention against no treatment (Example 3)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Author** | | |
|  |  | **Dunford (2000)** | **Bresnahan et al. (1999)** | **Ley et al. (2006)** |
| Outcome measure | Scale | ✓  Modified Conflict Tactics Scale | ✓  Battery of questionnaires including Conflict Tactics Scale | 🗶 |
|  | Reconviction or reoffending rates | ✓ | 🗶 | ✓ |
| Study design | | RCT | Quasi-experimental | Quasi-experimental |
| Setting | | Army | Offenders institution | Offenders institution |
| Duration of follow-up (days) | | 365 | 56 | 365 |
| Did the intervention result in a significant reduction in violence? | | 🗶 | 🗶 | ✓ |

RCT = randomised controlled trial

# Discussion, Limitations and Conclusion

## Discussion

The combined data from these two large reviews ([Hockenhull et al., 2012](#_ENREF_10); [Leitner et al., 2006](#_ENREF_14)) represent the most comprehensive dataset of available evidence on interventions to reduce violence. The scale of the dataset demonstrates the extensive and broad range of research that has been conducted in the area and the potential for a large collective evidence base, however, at the same time it has provided an opportunity to examine the scale and nature of heterogeneity amongst included studies. The results presented here confirm those found by previous reviews ([Brooks-Gordon et al., 2006](#_ENREF_3); [Duncan et al., 2006](#_ENREF_6); [Hassiotis & Hall, 2004](#_ENREF_9); [Huf et al., 2004](#_ENREF_11); [Smedslund et al., 2007](#_ENREF_21)) that the data are heterogeneous and that meaningful meta-analysis is not possible. Furthermore, the findings indicate that, even with the ability to perform a meta-analysis, firm conclusions cannot be drawn in some cases concerning the best treatment because the evidence is conflicting. This is well illustrated by Example 2; despite fulfilling the same stringent PICO criteria, two CBT-based interventions reported significant findings whereas two did not.

It could be argued that violence, broadly defined, is an inherently heterogeneous phenomenon and the heterogeneity of the literature merely reflects this. In other words, one size of study does not fit all sizes of violence. The wide-ranging subject area means that there are innumerable PICO tables that could be constructed. The population groups are in themselves heterogeneous and the interventions and comparators are frequently complex in their nature, once again preventing a homogenous dataset. Even drug trials which are highly homogenous in terms of the intervention can turn out to be heterogeneous in many other ways, such as the outcome measures used which, as this paper has shown, can vary considerably between studies. This, alongside the large number of confounding variables that would need to be controlled for in many of these studies, means that the subject area does not lend itself to a homogenous dataset.

Furthermore, the lack of a national or international funded programmatic approach to the problem (apart from a number of drug trials) has left the global research effort to individual practitioners and researchers who inevitably will design studies relevant only to their local population and problems. This leads to an over-reliance on relatively small convenience samples, audit data and retrospective designs. Many of these interventions are locally developed or instigated and are “unique”, causing additional heterogeneity. Pharmacological studies are frequently funded by pharmaceutical companies with the resources to conduct large scale studies employing methodological and statistical expertise. However, heterogeneity between studies of pharmaceutical products’ efficacy is still evident, as demonstrated in Example 1. Furthermore, some pharmaceutical companies have been known to misreport data for their own benefit ([Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008](#_ENREF_22)), indicating that published studies of pharmaceuticals showing positive effects are only a sample of those conducted and potentially also influencing the degree of heterogeneity between different pharmaceutical studies. The challenge is even greater in the field of psychosocial interventions as the same resources are rarely available for developers of psychological or environmental interventions.

To increase homogeneity of studies future studies need to be built on past research, utilizing similar participants, interventions, and particularly outcome measures wherever possible. The best way to reduce heterogeneity is through replication of high-quality studies. This replication does not have to be limited to exact replication but can be an extension of a previous study. Though convenience sample/audit type studies are laudable, promising results should be followed up and further tested using higher quality designs. By designing new studies to reflect, as closely as possible, previous studies using as similar PICO values as possible, it may be feasible to synthesize data from sub-groups. This change in the design of future studies would be helped by programme grants and funding streams prioritizing studies that aim to homogenize research in the field wherever possible.

Despite the heterogeneity evident in published studies investigating the efficacy of interventions to reduce violence in criminal or mental health populations, there have undoubtedly been many meaningful results and the effectiveness of interventions should not be underestimated. However, it would appear from this paper that unless more homogeneity is introduced into the literature, systematic review methodology is not a suitable method of evidence synthesis for the violence literature. Instead, a realist synthesis approach, as proposed by Pawson ([2006](#_ENREF_18)), may be more appropriate.

Realist synthesis incorporates theory-building in order to hypothesize and test relationships between contexts, mechanisms and outcomes, with the aim of understanding what works, for whom, and in what circumstances. This approach to evidence synthesis would seem highly appropriate in the area of violence research where interventions are complex both in the terms of the intervention delivered but also the populations treated and the outcomes used to measure effectiveness. It should, therefore, be considered alongside more ‘traditional’ systematic review methodology.

## Limitations

While based on an extensive database of intervention studies, informed by two large systematic reviews ([Hockenhull et al., 2012](#_ENREF_10); [Leitner et al., 2006](#_ENREF_14)), the extent of the data base meant that pragmatic decisions had to be made regarding sub-groups analyses conducted. This paper therefore only explored three such examples, albeit diverse in terms of population and intervention studied. It is possible that other examples may have produced more homogenous groups than those reported on in this paper, but an inspection of the data does not suggest this is likely.

An additional limitation is that due to the scale of the reviews and the considerable amount of work involved in their execution and publication, there is now an additional five years of publications (2008-2013) awaiting review. It is possible that evidence available from the last five years is more homogenous than that studied in this paper. However, updating the review to incorporate new literature is beyond the scope of this paper.

## Conclusions

Despite drawing from a pool of 496 studies, only one of the three chosen examples produced a group of studies suitable for meta-analysis and synthesis. However, even the results of this meta-analysis were too heterogeneous to meaningfully inform practice. This illustrates the heterogeneity of the research on violence and the difficulty presented to those who would like to use the current literature to inform practice and policy. Due to this heterogeneity, and, therefore, the lack of robustness of any synthesis it is not possible to draw any firm conclusions regarding what specific interventions effectively reduce violence. For firm conclusions to be possible, future research needs to either become more homogenous by focusing on specific high-priority topics or by adopting new approaches to evidence synthesis which can contain the complexity.

Thorough systematic reviews consolidate the best available evidence for an intervention and facilitate evidence-based practice. For results to be meaningful and robust, studies need to be as similar as possible. Results from these reviews show that the violence research outcomes literature is extensive but remains heterogeneous. This is often a result of the nature of the research but changes can be made so that future studies are more homogenous, thereby enabling meaningful synthesis of evidence and robust results. Where changes are not possible, such as in the evaluation of the effectiveness of complex interventions, a realist synthesis approach may provide an alternative and complementary method for clinicians and policy makers to inform clinical practice.

# Acknowledgements

The authors would like to thank Rachel Flentje, Beverley Quinn, Sarah Jones, Yenal Dundar and Jan Atkinson for their assistance in the conduct of the two previous systematic reviews which form the basis of this paper.

# References

Bhana, N., Foster, R., Olney, R., & Plosker, G. (2001). Olanzapine: an updated review of its use in the management of schizophrenia. *Drugs, 61*, 111-161.

Bresnahan, S. P. (1999). Men who are violent against their partners: A study evaluating the effectiveness of a domestic violence treatment program in a correctional setting. *Diss Abstr Int B Sci Eng, 60*(3-B).

Brooks-Gordon, B., Bilby, C., & Wells, H. (2006). A systematic review of psychological interventions for sexual offenders I: randomised control trials. *J Foren Psychi Psych, 17*(3), 442-466.

Citrome, L., Volavka, J., Czobor, P., Sheitman, B., Lindenmayer, J. P., McEvoy, J., et al. (2001). Effects of clozapine, olanzapine, risperidone, and haloperidol on hostility among patients with schizophrenia. *Psychiatr. Serv., 52*(11), 1510-1514.

Czobor, P., Volavka, J., & Meibach, R. C. (1995). Effect of Risperidone on Hostility in Schizophrenia. *J. Clin. Psychopharmacol., 15*(4), 243-249.

Duncan, E. A. S., Nicol, M. M., Ager, A., & Dalgleish, L. (2006). *A systematic review of structured group interventions with mentally disordered offenders*: Criminal Behaviour and Mental Health Vol 16(4) 2006, 217-241.

Dunford, F. W. (2000). The San Diego Navy Experiment: An assessment of interventions for men who assault their wives. *J. Consult. Clin. Psychol., 68*(3), 468-476.

Friendship, C., Mann, R. E., & Beech, A. R. (2003). Evaluation of a national prison-based treatment program for sexual offenders in England and Wales. *J Interpers Violence, 18*(7), 744-759.

Hassiotis, A., & Hall, I. (2004). Behavioural and cognitive-behavioural interventions for outwardly-directed aggressive behaviour in people with learning disabilities. *Cochrane Database Syst Rev*(4), CD003406.

Hockenhull, J., Whittington, R., Leitner, M., Barr, W., McGuire, J., Cherry, M., et al. (2012). Systematic review of intervention strategies for the prevention, treatment and management of violent behaviour by adults in contact with forensic mental health services or the criminal justice system. *Health Technol Assess, 16*(3).

Huf, G., Alexander, J., & Allen, M. H. (2004). Haloperidol plus promethazine for psychosis induced aggression. *Cochrane Database Syst Rev, 4*.

Kane, J. M., Marder, S. R., Schooler, N. R., Wirshing, W. C., Umbricht, D., Baker, R. W., et al. (2001). Clozapine and haloperidol in moderately refractory schizophrenia: a 6-month randomized and double-blind comparison. *Arch Gen Psychiatry, 58*(10), 965-972.

Krug, E., Dahlberg, L., Mercy, J., Zwi, A., & Lozano, R. (2002). *World report on violence and health*. Geneva: World Health Organization.

Leitner, M., Barr, W., McGuire, J., Jones, S., & Whittington, R. (2006). *Systematic review of prevention strategies for the forensic mental health population at high risk of engaging in violent behaviour. Final report to National Forensic Mental Health R&D Programme.* . Liverpool: University of Liverpool.

Ley, L. F. (2006). A study of domestic violence recidivism following treatment among incarcerated men who batter. *Diss Abstr Int B Sci Eng*.

McGrath, R. J., Hoke, S. E., & Vojtisek, J. E. (1998). Cognitive-behavioral treatment of sex offenders - A treatment comparison and long-term follow-up study. *Crim Justice Behav, 25*(2), 203-225.

Nicholaichuk, T., Gordon, A., Gu, D., & Wong, S. (2000). Outcome of an institutional sexual offender treatment program: A comparison between treated and matched untreated offenders. *Sex. Abuse, 12*(2), 139-153.

Pawson, R. (2006). *Evidence-Based Policy: A Realist Perspective*: Sage Publications.

Schmucker, M., & Losel, F. (2008). Does sexual offender treatment work? A systematic review of outcome evaluations. *Psicothema, 20*(1), 10-19.

Schweitzer, R., & Dwyer, J. (2003). Sex crime recidivism: evaluation of a sexual offender treatment program. *J Interpers Violence, 18*(11), 1292-1310.

Smedslund, G., Dalsbo, T. K., Steiro, A. K., Winsvold, A., & Clench-Aas, J. (2007). Cognitive behavioural therapy for men who physically abuse their female partner. *Cochrane Database Syst Rev*(3), CD006048.

Turner, E., Matthews, A., Linardatos, E., Tell, R., & Rosenthal, R. (2008). Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. *N. Engl. J. Med., 358*(3), 252-260.

U.S. National Library of Medicine. (2009). Medical Subject Headings (MeSH). Retrieved January, 2012, from <http://www.ncbi.nlm.nih.gov/mesh?term=evidence%20based%20practice>

World Health Organisation. (2013). Cause specific mortality: Regional estimates for 2000-2011. Retrieved July, 2013, from <http://www.who.int/healthinfo/global_burden_disease/estimates_regional/en/index.html>