MANDATORY ARREST FOR MISDEMEANOR DOMESTIC VIOLENCE EFFECTS ON REPEAT OFFENDING: PROTOCOL

Barak Ariel, PhD*  
Institute of Criminology  
University of Cambridge  
Sidgwick Avenue Cambridge  
CB3 9DA UK  
E-mail: ba285@cam.ac.uk  
Phone: +44(0)1223 767378  
Fax: +44(0)1223 768416

Professor Lawrence W. Sherman, PhD  
Institute of Criminology  
University of Cambridge  
Sidgwick Avenue Cambridge  
CB3 9DA UK  
E-mail: ls434@cam.ac.uk  
Phone: +44(0)1223 335360  
Fax: +44(0)1223 768416

* Lead/Contact reviewer

Latest Revision: 10 June 2012

Publication date: 1 September 2012

No Support Granted for this Review

Potential Conflicts of Interests: none known
1. BACKGROUND

In 1984, the Minneapolis domestic violence arrest experiment (Sherman and Berk, 1984), reported that arrest in misdemeanor domestic assault cases reduced reoffending substantially in comparison to two informal alternatives (mediation or separation). The evidence from this experiment strongly supported the hypothesis of specific deterrence theory, and these findings thus culminated in a rich academic discussion (Mills, 1998; Schmidt & Sherman, 1996; Tolman & Weisz, 1995) as well as elaborate public attention (National Institute of Justice, 2001). In many ways, this experiment paved the way to mandatory arrest policy in domestic violence cases in numerous countries around the western world.

Subsequent trials funded by the National Institute of Justice (commonly known as the Spousal Assault Replication Program (SARP)), however, failed to replicate the Minneapolis findings, reporting inconsistencies in the direction and impact of arrest in domestic violence cases (Berk et al., 1992a; Dunford et al., 1990a: 1990b; Hirschel et al., 1992; Pate and Hamilton, 1992; Sherman, 1992a).

Some systematic reviews of the NIJ studies were conducted (e.g., Garner et al., 1995; Maxwell et al., 2002; Sugerman & Boney, 2000). Across the completed replication studies, Maxwell (1998) and Maxwell et al. (2002) have reported the results of case-level analysis of the five replication studies using available offender and victim case-level data. They produced an overall effect (odd-ratios) for the arrest versus no arrest treatment comparison for the prevalence, frequency, and the time-to failure rates for all male offenders (n=4,032) and for the male offenders whose victims were interviewed after the incident and could report failure data (n=3,147). To account for differences in study designs and sample demographics, Maxwell et al. (2002) also weighted their pooled, cross-site effects by simultaneously controlling for the study sites and within study follow-up lengths, and several offender characteristics. Maxwell, et al. (2002) reports a modest yet consistent significant cross-site preventive effect of arrest on the rate of intimate partner violence re-victimization, using pooled analyses of the five replication studies.

Using data produced by Garner et al. (1995), Sugerman and Boney (2000, p. 67) reported that the average effect size for the arrest versus the non-arrest comparison on official recidivism rates was not significant, nor was it variable across the six sites. However, they do report a significant deterrent effect for arrest on the offenders’ rate of subsequent re-victimization as reported by the victims (p. 68). In the context of these victim-reported data, they also reported that the “variation among these six effect sizes

---

1 We wish to thank Campbell Collaboration’s anonymous reviewers and the Editor for their insightful commentary which we have incorporated in the text below.
may be accounted for by chance” (p. 68). Thus, this more traditional meta-analysis of published results produced conclusions that are substantively consistent to those articulated in Maxwell, et al.’s (2002) pooled analysis of case-level data from nearly the same studies. More precisely, both refereed studies found non-significant results for arrest across several re-arrest rates, but significant reductions in the re-victimization rate. They both also found that the results (e.g., effect sizes) replicated each other across the applicable studies. The only marginal difference produced by these two symmetric reviews was the degree to which assigning arrest influenced the rate of re-victimization. Maxwell, et al. (2002) reported an average reduction of about 25% in the prevalence of re-victimization attributed to assigning arrest while Sugerman and Boney (2002) reported finding just a 5% reduction due to arrest. This difference could be attributed to either the differences in the samples (Maxwell, et al. selected cases from just the five replication studies while Sugerman and Boney used effect sizes from all six completed studies), or because of the difference in their approaches to producing their synthesis (pooled data with covariates vs. meta-analyses of bivariate published summary results).

However, the available evidence from these trials and similar studies was not assessed under a full meta-analytic procedure within a Campbell review, measuring mean effect sizes and correlations between study features and effect sizes. Specifically, the reviews have not demonstrated across trials the relative magnitude of the difference between arrests and no-arrests, which can arguably provide a more informative assessment of arrest versus alternative-to-arrest police policies (although we take under consideration the possibility that the results might be equivocal). Neither Sugerman and Boney (2002) or Maxwell, et al. (2002) reported sub-treatment effects since both reviews chose to collapse the many non-arrest treatment groups into just one control group. Therefore, this review could add information about whether the arrest effects produced by these earlier syntheses varied across alternative control groups.

Lastly, previous studies and available reviews have not examined interaction effects with social bonds. Under the stake-in-conformity hypothesis (e.g., Toby, 1957; Sherman et al., 1992), it may be the case that there is an interaction effect of arrest with certain demographic characteristics of the suspect. Sherman and others (Sherman and Smith, 1992; Pate and Hamilton, 1992, and Berk et al., 1992b) have all reported a differential and interactive effect of arrest with the employment status of the suspect (as measured by victim interviews). Sherman (1992a) also reports similar but weaker interaction with marriage. Based on these earlier, less rigorous analyses, we propose to conduct moderator analysis of the available evidence in order to test this hypothesis. We hypothesize that that arrest deters employed
suspects from committing additional offenses, but increases subsequent domestic assault by unemployed suspects. We will also test the effect of marriage, education and age as moderators of arrest effects on repeat domestic violence; however the scope of the additional subgroup analyses largely depends on the data available in the original reports.

2. OBJECTIVES

The objective of this review is to systematically review and synthesize credible evidence on the effectiveness of arrest policy for misdemeanor domestic violence on repeat offending. Of interest are the effects on future reoffending in terms of same-victim violence and different-victim violence by the offender, and the moderating effect of stakes in conformity. We further plan to produce separate point estimates by specific control groups, which we believe will an improvement over the existing meta- and pooled-analyses. It is anticipated that this review will help inform policy-makers’ decisions regarding mandatory arrest policy. Many jurisdictions have already adopted and continue to enforce such policy and a critical examination of the existing evidence is warranted.

3. METHODOLOGY

3.1 CRITERIA FOR CONSIDERING STUDIES FOR THE REVIEW

3.1.1 TYPES OF INTERVENTION

The intervention of interest is mandatory arrest for misdemeanor domestic abuse. Mandatory arrest is defined as the legal duty of police to make an arrest if the officer has reason to believe a domestic violence act has been committed. Although different jurisdictions have different requirements to determine qualifying factors for mandatory arrest, they include at least the immediate physical removal from the scene of the incident and the physical incapacitation of the offender for a relatively short period of time (between a few hours to a couple of days). Arrest may also be used in conjunction with other immediate treatments, such as restraining orders, commitment to avoid additional
offences, etc. Such studies will be included as well, however we will assess mixed treatments separately, in order to learn how arrest interacts with other treatments as well.

In order to compare the effect of arrest, we will look for studies that either did not apply any intervention in the comparison group, or – in the more likely event – any alternative to an arrest, such as mediation, separation, restorative justice, consultation, citation, counseling as “ticket” citation, or any other processing carried out by the police in misdemeanor domestic violence cases that does not result in the suspect being taken into police custody at a different location. Collectively, we call this comparison category any “alternative to an arrest”, though we are keen to assess different ‘control groups’ separately as well.

3.1.2 TYPES OF STUDIES

We will include the following study designs (we will follow Campbell guidelines of reporting results separately for these two designs):

3.1.2.1 Field randomized controlled trials/true experiments that randomly allocate participants to arrest condition and an alternative condition.

3.1.2.2 Studies must include at least one alternative-to-an-arrest employed as the control intervention. Whenever more than one type of comparison intervention is tested, separate point estimates will be used for each type.

3.1.2.3 Quasi-experimental designs that include a control group and baseline assessment of comparability. While we recognize that the strength of evidence from these studies is generally, but not always, weaker than from true experiments, we are still interested in capturing the relationship between arrest and recidivism found in quasi-experimental designs so that we can fully summarize the existing evidence in this area. We therefore will look to identify:

3.1.2.3.1 An identifiable comparison group that does not receive arrest. This may be designed based on historical comparison group design.
Excluded comparison groups will include those who were ineligible for arrest (e.g., mental defects).

3.1.2.3.2 Studies that include baseline assessment of the comparability of arrest and comparison conditions. This means that we will assess the comparability of the study conditions and observed differences on several variables that may be associated with future criminal behavior and risk behaviors, such as critical demographic variables (i.e., age, employment, and marital status) as well as outcome measures at baseline.

3.1.3 TYPES OF PARTICIPANTS

3.1.3.1 Eligible studies will be based on samples of individuals (offenders) involved in a domestic abuse case. These may be both male and female participants, juvenile and adult participants, where participants of one of the treatment groups were arrested for misdemeanor domestic abuse in the presence of the police, shortly after a domestic abuse incident was reported to the police. In other words, only studies in which the offender was present at the scene upon arrival of police will be included in the review (cf. Dunford et al., 1990a).

3.1.3.2 The nature of misdemeanor domestic abuse suspects may vary in different jurisdictions. Collectively, these are cases in which one family member commits or threatens violence against another family member or household member - although we are specifically interested in violence between couples. These couples include any form of intimate partnerships, including spouses, roommates, housemates, dating partners, and same-sex partners.

---

2 The term “misdemeanor violence” mean a criminal offence that is punished less severely than felonies (e.g., punishable with incarceration for one year or less) but more severely than administrative infractions. Under American Federal or State law, for example, misdemeanor domestic violence “has a element, the use of, or attempted use of physical force, or the threatened use of a deadly weapon. It is an offense which was committed by a current or former spouse, parent, or guardian of the victim, by a person with whom the victim shares a child in common, by a person who is cohabiting with or has cohabited with the victim as a spouse, parent, or guardian, or by a person similarly situated to a spouse, parent, or guardian of the victim” (Title 18, United States Code, Section 922(g)(9); see also § 925(a)(1). In the review, we will separate between both same-victim and any-victim crimes, as more fully dealt with in 2.1.4.1 below.
3.1.4 TYPES OF OUTCOME MEASURES

3.1.4.1 The primary outcome of interest in this systematic review is criminal behavior following the arrest for domestic violence, against the same victim. Outcome data may comprise of official records such as arrests, charges or re-convictions for a new offense; or victim self-reported victimization made by the same-victim of the said arrest. These various outcome measures will be analyzed separately, as discussed below.

3.1.4.2 The secondly outcome of interest in this systematic review is the same as the primary, except it includes recidivism data against any victim.

3.1.4.3 If outcome measures are not reported separately or are not clearly defined as such, we will assume they are meant to pertain to any-victim crimes – although a cursory review of the literature indicates that such specificity is usually reported.

3.1.5 SETTINGS AND TIMEFRAME

We will not exclude studies on the basis of language or geography. We will work with our international contacts to learn which countries are likely to have used and evaluated the effect of arrest in misdemeanor domestic violence so that we can target foreign language and location searches appropriately. Studies using data collected from 1970 onwards will be included. The rationale for this timeframe is twofold: first, it is unlikely that any older studies with eligible research designs exist; and second, the social context is markedly different. For example, the police culture changed extensively between the 1980s and 1990s in its views on domestic affairs. Furthermore, the structure of the criminal justice system has changed considerably since the 1960s, which would potentially make the coding of domestic violence-related offense types difficult.

Studies that do not meet one or more of the above criteria will be excluded from the review altogether, or just from the relevant portion of the review. Both reviewers will independently
screen the full text of studies and recommend whether to include them in the review. Disagreements will be resolved upon discussion and if no resolution is achieved, the study will not be included in the review. Each excluded study will be listed along with a reason for its exclusion.

3.2 SEARCH METHODS FOR IDENTIFICATION OF STUDIES

3.2.1 The search strategy will include the following sources:

3.2.1.1 Searches for NIJ-funded experiments on the impact of arrest in domestic abuse incidents;

3.2.1.2 Extensive search of online databases (see section 2.4 below);

3.2.1.3 Searches of narrative and empirical reviews of literature that examined the effectiveness of arrest in preventing subsequent domestic abuse.

3.2.1.4 Search for literature reviews on the relationship between arrest and domestic abuse;

3.2.1.5 Searches of bibliographies on the effectiveness of arrest in preventing domestic abuse;

3.2.1.6 Registers of randomized controlled trials: the Registry of Randomized Experiments in Criminal Sanctions, 1950 – 1983 (Weisburd, Sherman and Petrosino 1990) and the Social Psychological, Educational and Criminological Trials Register (SPECTR) developed by the Cochrane Centre;

3.2.1.7 As studies will be located, their references will be examined for details on other relevant studies. These will then be examined with accompanying notes being made to explain where the document was originally cited;

3.2.1.8 Contact with key researchers in this field.

3.2.2 Each title and abstract will then be screened to establish if it meets the criteria established in Section 3.1 above. These studies will be assessed using the checklist attached hereto as Appendix I. The checklist form will be completed for each of the studies.
3.2.3 No limitations are made on the nature of publication (i.e., published or unpublished material and ‘gray literature’). See Section 3.6.2 below in relation to publication bias analysis.

3.3 SEARCH TERMS

Three categories of keywords were developed for this search. The intention of separating the terms in this manner is to include all the potentially relevant results, while simultaneously excluding the large bodies of literature on domestic violence from non-criminological disciplines. These sets of keywords will be combined with a Boolean “AND”.

3.3.1 Policy of Interest

[“ARREST*” or “CRIME*” “OFFEND*” or “BATTER*” or “SUSPECT”] and [“DOMESTIC VIOLENCE” or “DOMESTIC ABUSE*” or “DOMESTIC ASSAULT*” or “MARITAL VIOLENCE” or “BATTERED WOMEN” or “BATTERED PARTNERS” or “SPOUSAL ABUSE” or “WIFE BEATING*” or “INTIMATE PARTNER VIOLENCE” or “FAMILY VIOLENCE” (however violence solely against non-partner members, such as teen on parents or juveniles, will be excluded);

3.3.2 Outcomes

CRIM* or DELINQUEN* or ARREST* or DETAIN* or DETENTION or “CALL* FOR SERVICE*” or OFFEND* or VIOLEN* or ASSAULT or FIGHT* or RE-ARREST or RECIDI* or “DETER” or “CONFLICT TACTICS SCALE” or “CTS” or “CTS2” or “CTSPC” or “RECONVICTION”

3.3.3 Research Design

“EXPERIMENT” or “QUASI-EXPERIMENT” or “RANDOMIZED CONTROLLED TRIAL” or “RCT” or “RANDOM ASSIGNMENT” or “FIELD EXPERIMENT” or “EVALUATION AND COMPARISON” or “EVALUATION AND CONTROL”

Key terms will also be used in conjunction with “DISSERTATION” or “THESIS” to locate such works as well.
3.4 **ELECTRONIC SOURCES**

3.4.1 The databases listed in Appendix III will be searched for eligible studies (list of databases appears in alphabetical order), followed by a search in Google Scholar.

3.4.2 We will also collect subject-level data from the *Interuniversity Consortium for Political and Social Research* (ICPSR), which stores the collected data from at least six known replication studies funded by the National Institute of Justice with the aim of replicating the Minnesota Domestic Violence Experiment (Sherman & Berk, 1984). This will enable us to download the six replication studies from ICPSR, each containing multiple data files that will need merging, to produce the stakes in conformity models using criminal history and victim interview data and ultimately conduct the analyses listed in 3.9.2 below:

3.4.3 The bibliographies of relevant articles from specialized journals (e.g., *Journal of Interpersonal Violence* and *Journal of Family Violence*) will be reviewed as well.

3.5 **CODING OF STUDIES**

The two independent reviewers will extract information from full-text versions of eligible studies using the coding protocol (see Appendix II). If both reviewers agree on the rating results of each article and the coding of the data from each article, then the data will be entered into SPSS and Comprehensive Meta-Analysis 2.0.

The following issues will be dealt with in the coding process:

3.5.1.1 Effect-sizes for all available time periods will be coded (i.e., 6-, 12-, 18 and 24-months follow-up period).

3.5.1.2 The outcome measure will be reported separately for two main categories: frequency data gathered from official records (police arrest or offense reports) and prevalence data from victim surveys (initial domestic violence victims).

3.5.1.3 Because of our aim to assess whether the effect of arrest was conditional upon “stakes in conformity” (Toby, 1957), any social “stakes measures” for the offender will be
codes as well, including other potential moderators such as age, gender and education levels.

It is likely that information on these indicators will appear in the unofficial reports, such as interviews conducted with either the offenders or their victims. At the same time, this source of information is problematic due to likely poor response rate. Therefore, any study in which the response rate is less than 60% will be excluded from that portion of the review.

3.6 DATA COLLECTION AND ANALYSIS

3.6.1 ASSESSMENT OF RISK OF BIAS IN INCLUDED STUDIES

The extent to which we can draw conclusions about the effect of arrest in domestic violence cases depends on the validity of the outcomes of the primary studies. We are particularly concerned about internal and external validities, given the arguably non-comparability of studies. The reliability of the results may also be at risk, given the methodologies used in primary studies, should low-level studies be included as well.

We plan on using critical assessment for various risk domains in a checklist format, proposed by Juni (2001). This list appears in a tabular format in Appendix IV and it contains five types of biases:

3.6.1.1 Selection bias (i.e., systematic differences between baseline variables that define the groups before the arrest);

3.6.1.2 Attrition bias (i.e., systematic differences between the groups in withdrawals or exclusions of participants from the results of a trial; in this context, we specifically mean systematic differences between participants who were arrested as assigned and those that were not arrested though assigned such treatment);

3.6.1.3 Performance bias (i.e., systematic exposure to factors other than arrest, specifically in the comparison groups);

3.6.1.4 Detection bias (i.e., systemic measurement differences); and
3.6.1.5  Reporting bias (e.g., selective outcome reporting).

Some items are objective and quite apparent (e.g., the participants were selected and allocated in non-random procedures, precise exclusion criteria were not always used in the selection of the participants, and studies have not incorporated power calculations), and some are subjective (e.g., can the design address the studied question in a comprehensive way?). We will take a robust approach by stating whether within each study there is a “low” “medium” or “high” level of bias on every risk domain, which we will score independently. We will then review these scores to obtain a measurement of each bias across the studies, in order to assess whether their plausible impact on the outcomes. We will also record the source of each bias as well.

3.6.2  ASSESSMENT OF PUBLICATION BIAS

Publication bias can lead to systematic bias in our review. We will estimate the reporting bias in published versus unpublished works using funnel plots (Rothstein, Sutton and Borenstein, 2005). Funnel plots can be used to assess whether a systematic review is likely to be vulnerable to publication bias, by plotting arrest treatment effect (i.e. mean difference between intervention group and control) against the inverse of the variance or the sample size.

However, we will only explore this option should enough studies meet our eligibility criteria.

3.7  DESCRIPTION OF METHODS USED IN PRIMARY RESEARCH

Very few studies randomly assign domestic violence offenders to either arrest or an alternative-to-arrest groups; a cursory review of the literature suggests that the NIJ replication (e.g., Dunford 1990; Maxwell 1998; Schmidt and Sherman 1993) and the Sherman and Berk (1984) Domestic Violence Minneapolis experiment are the leading if not the only experiments in this field. Some studies have had after-only measures of repeat victimization, of those victims whose partners were arrested compared to those whose partners were not arrests; such studies controlled for baseline equality statistically (e.g., Cho and Wilke
2010). Most studies looked at official records as the main outcome variable (e.g., Hirschel et al. 1992), where others looked at victim interviews following random assignment to arrest and control conditions (e.g., Jolin et al. 1999).

3.8 **CRITERIA FOR DETERMINATION OF INDEPENDENT FINDINGS**

3.8.1 **Combining Multiple Outcomes**

3.8.1.1 There are likely several independent outcomes (e.g., arrest and reconviction) or time-points within each study (e.g., arrest within 6 months or 12 months), and we will report and synthesize each separately (for example, a separate meta-analysis for arrest data and a separate for reconviction data). We hope that sufficient data will be available to cluster the available information within such homogeneous outcomes.

3.8.1.2 In the case of studies the measure more than one alternative to an arrest, these will be first be collapsed into a single categorical condition of “alternative-to-an-arrest”, but will also be analyzed independently, based on the number of studies available under each category, as described below.

3.8.2 **Multiple Subgroups**

3.8.2.1 We also plan to analyze the following subgroups within studies, where possible. First, if sufficient data are reported in the primary studies, we will explore the impact of potential covariates on the outcomes, using random effects meta-regression or analogue-to-the-ANOVA moderators’ analysis with SPSS META macro (Professor Wilson, 2003). We thus want to be able to look into the various treatment components as well as extraneous elements that may influence the results, such as treatment components (long versus short arrest; with or without additional treatments) and methodological quality of the study (e.g., experimental, non-experimental).

---

3.8.2.2 More importantly, as we emphasized earlier, we place particular importance on the conditional effect of arrest on stakes in conformity - primarily marriage and employment, however other features of the participants are of interest as well (age, gender, education, and ethnicity).

3.9 MEASURES OF TREATMENT EFFECT

The primary outcome measure will be the frequency of official police records of repeat arrests or reports of domestic violence with the same offender against the same victim. The secondary outcome measure will be the same as the primary, except that it will include any additional non-domestic violence offending. The tertiary measure will be the victim-reported measures of reoffending. We elaborate on these processes below.

3.10 EFFECT SIZE CALCULATIONS AND DATA SYNTHESIS

3.10.1 The odds-ratio will be the effect size of choice for all outcomes of a dichotomous or binary nature, calculated by comparing arrest and alternative-to-an-arrest conditions (that is, our tertiary outcome of interest). The standardized mean difference (SMD) effect size in the form of Cohen’s $d$ (Cohen 1988) will be used for continuous or count data (that is, our primary and secondary outcomes of interest).

3.10.2 Outcomes will be meta-analyzed using traditional inverse-variance weighted meta-analysis if possible. In all cases, a random effects model will be assumed a priori.

3.10.3 Cochrane-Q and $I^2$ statistic will be used to measure for homogeneity and heterogeneity between the tests.

3.10.4 In order to adjust for baseline differences in quasi-experimental studies, we will use the difference-in-differences (post-test mean minus the post-test mean for each group) in the numerator of each effect-size (while the denominator continues to be the raw within-group standard deviations, not the standard deviation of differences).

---

4 $I^2$ is expressed as a percentage of the total variance in all the data, with 25%, 50% and 75% considered as low, moderate or high level of heterogeneity.
3.10.5 **Moderator Analysis.** We will then explore potential covariates that were previously hypothesized to moderate the effect of arrest on certain individuals. Through formal moderator analysis using linear multiple regression, we will examine the pattern of evidence for the conformity stakes and how they affect the relationship between arrest and recidivism (Toby 1957; Sherman 1992), in each subset of data. We will interpret the findings in light of the methodological quality and potential bias inherent in the study designs\(^5\).

### 3.11 TREATMENT OF QUALITATIVE RESEARCH

Qualitative research will not be included in this review or in the meta-analysis.

### 4. TIMEFRAME

We envisage completion of the review within 10 days of approval of this protocol.

### 5. PLANS FOR UPDATING THE REVIEW

The review will be updated on a three-year basis. As part of this update we will need to code any new studies identified and rerun the analyses.

### 6. ACKNOWLEDGEMENTS

We wish to thank the anonymous reviewers of this protocol for their valuable comments. We would particularly like to thank Professor David Wilson for his insightful suggestions.

### 7. STATEMENT CONCERNING CONFLICT OF INTERESTS

None.

---

\(^5\) Because they are present prior to randomization, the proposed moderator variables should be uncorrelated with treatment assignment and we do not envisage requiring any specialised handling given the ITT approach.
8. REFERENCES


9. APPENDICES

Appendix I  Eligibility Criteria Checklist

Appendix II  Coding Protocol

Appendix III  List of Database

Appendix IV  Assessment of Biases
APPENDIX I – ELIGIBILITY CRITERIA CHECKLIST

Are the following inclusion criteria present (please answer with “yes” or “no” or provide answers as instructed; If not clear, attempt to find answers from the author)?

a) The report deals with arrest versus alternative-to-an-arrest in misdemeanor domestic violence cases (hereinafter domestic violence = DV).

_______

b) The study incorporates at least one comparison group.

_______

c) The study has reported a post arrest measure of continued domestic abuse.

_______

d) The effectiveness of arrest was measured by comparing participants who received it (the experimental group) with participants who did not receive it (the control group).

_______

e) The study had some control of extraneous variables (establishing the prior equivalence of groups) by (i) randomization, or (ii) pretest measures of arrest in DV, or (iii) matching, or (iv) pretest measurement of risk factors or risk scores for DV.

_______

f) Data that can be converted into computable effect-sizes are reported

_______

FOR A STUDY TO BE ELIGIBLE, ALL CRITERIA MUST BE MET, AND DETAILS NEED TO BE PROVIDED IN THE CODING SHEET
**Appendix II - STUDY LEVEL CODING PROTOCOL**

Use one study level code sheet for each study. If multiple documents report on the results from the same study, identify one of the documents as primary and use its document ID as the StudyID below.

*Throughout, 999 = not reported, 1 = Yes, 2 = No*

<table>
<thead>
<tr>
<th>Date of coding</th>
<th>________________________________</th>
</tr>
</thead>
</table>

**A. GENERAL DETAILS**

**Identifying Information:**

1. Study (document) identifier                      StudyID _____
2. Cross reference document identifier              CrossRef1 _____
5. Coder's initials                                 SCoder _____
6. Date coded                                      Date ___ - ___ - ___

**General Study Information:**

7. Author                                          ________________________________
8. Funder (e.g., NIJ)                               ________________________________
9. Geographical Location of Study                   SLocale ________________________________
10. Geography (1=single site; 2=multiple sites; 9=cannot tell) Sites _____
11. Country                                        ________________________________
12. Date range for data collection                  StartDate ___ - ___ - ___         DoneDate: ___ - ___ - ___
13. Publication Type                                PubType _____
   1. Book                                          4. Book Chapter
   2. Journal (peer reviewed)                        5. Federal Gov't Report

14. Information on the Domestic Violence Arrest Policy:

| ________________________________ |
B. STUDY AIMS AND RATIONALE

16. What are the aims of this study? [aims]

17. Why was the study done at that point in time, in this context and with those participants? [why]

18. What are the research questions and/or hypotheses? [h1]

C. STUDY SAMPLE

19. Harmonic n (at random assignment stage) [HarmonicN]

20. N in each subgroup (at random assignment stage) [Nsubgroup]

21. Attrition rate [Attrition]

22. N receiving arrest treatment [Ntreatment]

23. N receiving alternative to an arrest treatment (control) (if more than one control group list all) [Ncontrol]

24. Place of residence (in each subgroup) [Residence]
25. Mean Age of offenders (in each subgroup)  [AgeOffend]

26. Mean Age of Victims (in each subgroup)  [AgeVictim]

27. Gender (in each subgroup)  [Gender]

28. Socio-economic status (in each subgroup)  [socio]

29. Ethnic/ religion background (in each subgroup)  [ethnic]

30. % married (in each subgroup)  [marriage]

31. % with high school diploma (in each subgroup)  [diploma]

32. % Employed (in each subgroup)  [employed]

33. Size of family (in each subgroup)  [family-size]

D. INTERVENTION

34. Operationalization of conditions:  [operational]

35. Duration of intervention:  [duration]

36. Follow up period (in months):  [followup]

37. Who provides the intervention?  [provider]

38. Was special training given to people providing the intervention?  [training]
39. Arrest measurement at baseline – measurement of central tendency [baseline_CT]

40. Arrest measurement at baseline – measurement of dispersion [baseline_D]

41. Non-Arrest treatment type: [nonarrest]

42. Number of unique treatment programs (for each subgroup): [ngroups]

43. Type and components of treatments: [treatment_types]
   1) Arrest (in experimental group) Yes = 1, No = 2
   2) Mediation (in control group) Yes = 1, No = 2
   3) Separation (in control group) Yes = 1, No = 2
   4) Consultation (in control group) Yes = 1, No = 2
   5) Any other alternative-to-an-arrest Yes = 1, No = 2

E. RESEARCH DESIGN

44. Design Type [DesignType]
   1) Historical Controls
   2) Interrupted times-series (with comparison series)
   3) Non-equivalent comparison group with pre-post test
   4) Equivalent comparison group with pre-post test
   5) Randomized controlled trial (true experiment)
   6) Other __________________

45. Number of measures prior to implementation of intervention [PreMeasur]

46. Number of measures following implementation of arrest [PostMeasur]

47. Baseline characteristics measures [Baseline]

48. Analysis type [Analysis]
   1) t-tests
   2) F-ratios
   3) Time-to-failure
   4) OLS regression
   5) Other regression
6) Simple comparison of rates
7) Other

49. Analysis adjusted for covariates

50. Type of random assignment (if relevant)
1) Simple
2) Blocked random assignment
3) trickle-flow
4) batch
5) minimization
999) not relevant

51. How were units allocated to experimental and control conditions?
1) Randomly
2) Haphazardly
3) Selection effect
999) not relevant

52. Were Control conditions comparable?
1) Yes
2) No
999) Unknown

53. Variables measured to establish matching or comparability

54. To what population can the results be generalized?

55. Is there a potential generalizability threat from overall attrition or cross over?
1) Yes
2) No
999) Unknown

56. Is there potential threat to internal validity from differential attrition?
1) Yes
2) No
999) Unknown
EFFECT LEVEL CODE SHEET

Code these items for each eligible study

57. Identifying Information:

1) Study (document) identifier ________________________ [StudyID]
2) Effect identifier ________________________ [ESID]
3) Coder's initials ________________________ [GrpCoder]
4) Date coded ________________________ [Date_coded]

58. Outcome Type:

58.1 Label for outcome: _________________________________ [label]
55.2 Primary outcome? [Primary]
   1) Same victim
   2) Any victim
   999) Unknown

55.3 Outcome type treatment [TxOutType]
   55.3.1 Domestic violence reports
   55.3.2 CTS
   55.3.3 Arrest
   55.3.4 Conviction

55.4 Outcome type control [CgOutType]
   55.4.1 Mediation
   55.4.2 Separation
   55.4.3 Consultation
   55.4.4 Any other alternative-to-an-arrest

55.5 Source of data [DataSrc]
   55.5.1 Official records
   55.5.2 Victim self-reports
   55.5.3 Both

59. Sample size information

59.1 Unit-of-analysis: _________________________________ [Unit-of-analysis]
59.2 Number of units, arrest condition: ______ [TxN]
59.3 Number of units, comparison condition 1 ______ [CgN1]
59.4 Number of units, comparison condition 2 ______ [CgN2]
59.5 Number of units, comparison condition 3 ______ [CgN3]
60. Follow up period

60.1 6 month post random assignment
60.2 12 month post random assignment
60.3 18 month post random assignment
60.4 24 month post random assignment

[follow-up1]

61. Effect Data

61.1 Baseline – arrest group mean (Same-Victim)  
61.2 Baseline – arrest group S.D. (Same-Victim)  
61.3 Baseline – arrest group mean (Any-Victim)  
61.4 Baseline – arrest group S.D. (Any-Victim)  
61.5 Baseline – arrest group mean (Victim-Report)  
61.6 Baseline – arrest group S.D. (Victim-Report)  
61.7 Baseline – comparison group1 mean (Same-Victim)  
61.8 Baseline – comparison group1 SD (Same-Victim)  
61.9 Baseline – comparison group1 mean (Any-Victim)  
61.10 Baseline – comparison group1 SD (Any-Victim)  
61.11 Baseline – comparison group1 mean (Victim-Report)  
61.12 Baseline – comparison group1 SD (Victim-Report)  
61.13 Baseline – comparison group2 mean (Same-Victim)  
61.14 Baseline – comparison group2 SD (Same-Victim)  
61.15 Baseline – comparison group2 mean (Any-Victim)  
61.16 Baseline – comparison group2 SD (Any-Victim)  
61.17 Baseline – comparison group2 mean (Victim-Report)  
61.18 Baseline – comparison group2 SD (Victim-Report)  
61.19 Intervention – arrest group mean (Same-Victim)  
61.20 Intervention – arrest group S.D. (Same-Victim)  
61.21 Intervention – arrest group mean (Any-Victim)  
61.22 Intervention – arrest group S.D. (Any-Victim)  
61.23 Intervention – arrest group mean (Victim-Report)  
61.24 Intervention – arrest group S.D. (Victim-Report)  
61.25 Intervention – comparison group1 mean (Same-Victim)  
61.26 Intervention – comparison group1 SD (Same-Victim)  
61.27 Intervention – comparison group1 mean (Any-Victim)  
61.28 Intervention – comparison group1 SD (Any-Victim)  
61.29 Intervention – comparison group1 mean (Victim-Report)  
61.30 Intervention – comparison group1 SD (Victim-Report)  
61.31 Intervention – comparison group2 mean (Same-Victim)  
61.32 Intervention – comparison group2 SD (Same-Victim)  
61.33 Intervention – comparison group2 mean (Any-Victim)  
61.34 Intervention – comparison group2 SD (Any-Victim)  
61.35 Intervention – comparison group2 mean (Victim-Report)  
61.36 Intervention – comparison group2 SD (Victim-Report)  
61.37 SE adjusted for covariates (1=yes; 0=no)  
61.38 Unstandardized coefficient from statistical model  
61.39 SE for unstandardized coefficient from statistical model
61.40 Type of statistical model

61.40.1 OLS regression
61.40.2 Other egression T-test
61.40.3 F-ratio
61.40.4 Simple comparison of rates
61.40.5 Survival
61.40.6 Other

62. Hand Calculated Effect Size

Using the Effect Size Calculator by David Wilson

62.1 Hand calculated d-type effect size (primary outcome) [PES_Hand1]
62.2 Hand calculated standard error of the d-type effect size (primary) [PES_Hand2]
62.3 Hand calculated d-type effect size (secondary outcome) [SES_Hand1]
62.4 Hand calculated standard error of the d-type effect size (secondary) [SES_Hand2]
62.5 Hand calculated odds-ratio effect size (tertiary outcome) [TES_Hand3]
62.6 Hand calculated odds-ratio standard error (tertiary outcome) [TES_Hand4]
62.7 Effect Size adjusted for covariates (1=yes; 0=no) [CovAdj]
62.8 Page of effect size data [ESPage]
62.9 Effect Size Calculator Computation Type
Appendix III - List of Electronic Databases

1. Google, Google Scholar, Google Books
2. Academic Search Premier
3. C2 SPECTR
4. CINCH
5. Criminal Justice Abstracts
6. Criminal Justice Periodical Index
7. ERIC
8. ESDS
9. Healthsource Nursing Academic Edition
10. Ingenta
11. International Bibliography of Social Sciences
12. Medline
13. NCJRS
14. Proquest Digital Dissertations
15. Psyc Articles
16. Psych Info
17. Raven Web of Knowledge
18. Science Direct Scopus
19. Social Sci Search
20. Social Science Citation Index
21. Social Work Abstracts
22. Sociological Abstracts
23. SOSIG Law
24. SOSIG Social and Political Science
25. Zetoc – Electronic Table of Content
26. Springerlink
27. Wiley Online Library
28. Sage Libraries
29. Gray Literature Database (http://lawlibrary.rutgers.edu/cj/gray/)
30. Journal of Interpersonal Violence
31. Journal of Family Violence
32. Westlaw UK and Westlaw UK
Appendix IV

ASSESSMENT OF BIASES

<table>
<thead>
<tr>
<th>Risk of Bias</th>
<th>Number of Studies characterized by the bias</th>
<th>Source / Cause of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attrition bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detection bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reporting bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection bias</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Approach for summary assessments of the risk of bias for each important outcome (across domains) within and across studies

<table>
<thead>
<tr>
<th>Risk of bias</th>
<th>Interpretation</th>
<th>Within a study</th>
<th>Across studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low risk of bias</td>
<td>Plausible bias unlikely to seriously alter the results</td>
<td>Low risk of bias for all key domains</td>
<td>Most information is from studies at low risk of bias</td>
</tr>
<tr>
<td>Unclear risk of bias</td>
<td>Plausible bias that raises some doubt about the results</td>
<td>Unclear risk of bias for one or more key domains</td>
<td>Most information is from studies at low or unclear risk of bias</td>
</tr>
<tr>
<td>High risk of bias</td>
<td>Plausible bias that seriously weakens confidence in the results</td>
<td>High risk of bias for one or more key domains</td>
<td>The proportion of information from studies at high risk of bias is sufficient to affect the interpretation of results</td>
</tr>
</tbody>
</table>