

Campbell Systematic Reviews

2013:10

First published: 01 November, 2013

Last updated: October, 2013

Search executed: 04 September, 2012

Restorative Justice Conferencing (RJC) Using Face-to-Face Meetings of Offenders and Victims: Effects on Offender Recidivism and Victim Satisfaction. A Systematic Review

Heather Strang, Lawrence W Sherman, Evan Mayo-Wilson, Daniel Woods, Barak Ariel.



THE CAMPBELL COLLABORATION

Title	Restorative Justice Conferencing (RJC) Using Face-to-Face Meetings of Offenders and Victims: Effects on Offender Recidivism and Victim Satisfaction. A Systematic Review.
Authors	Heather Strang ¹ , Lawrence W Sherman ¹ , Evan Mayo-Wilson ² , Daniel Woods ³ , Barak Ariel ⁴ . ¹ Jerry Lee Centre for Experimental Criminology, Institute of Criminology, University of Cambridge, UK ² University College, London, UK ³ Police Executive Research Forum, Washington, DC, USA ⁴ Institute of Criminology, University of Cambridge, UK
DOI	DOI: 10.4073/csr.2013.10
No. of pages	171
Citation	Strang H, Sherman LW, Mayo-Wilson E, Woods D, Ariel B. Restorative Justice Conferencing (RJC) Using Face-to-Face Meetings of Offenders and Victims: Effects on Offender Recidivism and Victim Satisfaction. A Systematic Review. <i>Campbell Systematic Reviews</i> 2013;9 DOI: 10.4073/csr.2013.10
Copyright	© Strang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.[delete if co-registered]
Contributions	Heather Strang and Lawrence W. Sherman contributed to the writing and revising of this Review. Daniel J. Woods and Barak Ariel contributed to the statistical analysis. The search strategy was carried out by Evan Mayo-Wilson. Heather Strang will be responsible for updating this review as additional evidence accumulates and as funding becomes available.
Editors for this review	Editor: David Wilson Managing editor: Charlotte Gill
Support/funding	This systematic review received funding from the Norwegian Knowledge Centre for the Health Sciences, the Jerry Lee Center for Criminology at the University of Pennsylvania, the Smith Richardson Foundation, and the Jerry Lee Centre for Experimental Criminology at the University of Cambridge.
Potential conflicts of interest	The authors have no monetary interest in the results of the review. None of the authors has conducted or published studies that would lead them to slant the evidence on restorative justice in a particular direction. The two senior authors have published studies showing restorative justice both increases crime and reduces it.
Corresponding author	Heather Strang, Ph.D., Jerry Lee Centre for Experimental Criminology, Institute of Criminology University of Cambridge Cambridge CB3 9DA United Kingdom Email: hs404@cam.ac.uk

Campbell Systematic Reviews

Editors-in-Chief Julia Littell, Bryn Mawr College, USA
Arild Bjørndal, The Centre for Child and Adolescent Mental Health, Eastern and Southern Norway & University of Oslo, Norway

Editors

Crime and Justice David B. Wilson, George Mason University, USA

Education Sandra Wilson, Vanderbilt University, USA

Social Welfare Nick Huband, University of Nottingham, UK
Geraldine Macdonald, Queen's University, UK & Cochrane Developmental, Psychosocial and Learning Problems Group

International Development Birte Snilstveit, 3ie, UK
Hugh Waddington, 3ie, UK

Managing Editor Karianne Thune Hammerstrøm, The Campbell Collaboration

Editorial Board

Crime and Justice David B. Wilson, George Mason University, USA
Martin Killias, University of Zurich, Switzerland

Education Paul Connolly, Queen's University, UK
Gary W. Ritter, University of Arkansas, USA

Social Welfare Aron Shlonsky, University of Toronto, Canada
Jane Barlow, University of Warwick, UK

International Development Peter Tugwell, University of Ottawa, Canada
Howard White, 3ie, India

Methods Therese Pigott, Loyola University, USA
Ian Shemilt, University of Cambridge, UK

The Campbell Collaboration (C2) was founded on the principle that systematic reviews on the effects of interventions will inform and help improve policy and services. C2 offers editorial and methodological support to review authors throughout the process of producing a systematic review. A number of C2's editors, librarians, methodologists and external peer-reviewers contribute.

The Campbell Collaboration
P.O. Box 7004 St. Olavs plass
0130 Oslo, Norway
www.campbellcollaboration.org

Table of contents

Synopsis/Abstract	2
Objective	2
Conclusions	2
Executive Summary	3
Background	3
Objectives	3
Search Strategy	3
Selection Criteria	3
Data Collection And Analysis	4
Results	4
Authors' Conclusions	5
Role of the Authors in Reviewed Studies	6
1 Background	7
1.1 Definition of RJC	8
1.2 Theoretical Basis	9
2 Objectives	11
3 Methodology	12
3.1 Criteria for Considering Studies for This Review	12
3.2 Search Strategy For Identification of Studies	15
3.3 Selection of Studies	17
3.4 Data Management And Extraction	18
3.5 Outcome Measures	18
3.6 Effect Size Estimates And Moderator Analyses	18
4 Results	20
4.1 Description Of Studies	20
4.2 Assessment Of Methodological Quality Of Included Studies	21
4.3 Meta-Analysis Of Repeat Offending	24
4.4 Moderator Analyses	26
4.5 Sensitivity Analyses	29

4.6	Evaluator Effects.	29
4.7	Victim Impact Measures: Narrative Findings From The Ten Included Studies	33
4.8	Cost Effectiveness	44
5	Authors' Conclusions	47
5.1	Implications For Practice	47
5.2	Implications For Research	48
6	Acknowledgements	50
7	Potential Conflict Of Interest	51
8	Contact Details For Co-Reviewers	52
9	References To Tests Considered For Inclusion	53
10	Other References	55
11	Appendix A	59
11.1	Studies Excluded From The Review:	59

Synopsis/Abstract

OBJECTIVE

This systematic review examines the effects of the subset of restorative justice programs that has been tested most extensively: a face-to-face Restorative Justice Conference (RJC) “that brings together offenders, their victims, and their respective kin and communities, in order to decide what the offender should do to repair the harm that a crime has caused” (Sherman and Strang, 2012: 216). The Review investigates the effects of RJs on offenders’ subsequent convictions (or in one case arrests) for crime, and on several measures of victim impact. The review considers only randomized controlled trials in which victim and offenders consented to meet prior to random assignment, the analysis of which was based on the results of an “intention-to-treat” analysis. A total of ten experiments with recidivism outcomes were found that met the eligibility criteria, all of which also had at least one victim impact measure.

CONCLUSIONS

Our synthesis of these experiments shows that, on average, RJs cause a modest but highly cost-effective reduction in repeat offending, with substantial benefits for victims. A cost-effectiveness estimate for the seven United Kingdom (UK) experiments found a ratio of 8 times more benefit in costs of crimes prevented than the cost of delivering RJs.

Executive Summary

BACKGROUND

“Restorative justice” is a concept denoting a wide range of justice practices with common values, but widely varying procedures (Braithwaite, 2002). These values encourage offenders to take responsibility for their actions and to repair the harms they have caused, usually (although not always) in communication with their personal victims. This review focuses on the subset of restorative justice procedures that has been tested most carefully and extensively: face-to-face restorative justice conferencing (RJC). In these conferences, victims and offenders involved in a crime meet in the presence of a trained facilitator with their families and friends or others affected by the crime, to discuss and resolve the offense and its consequences.

OBJECTIVES

The reviewers sought to assess the effect of face-to-face restorative justice conferencing on repeat offending and on available measures of victim impact.

SEARCH STRATEGY

To identify studies eligible for inclusion in the review, 15 electronic databases were searched, including: Criminal Justice Abstracts, Dissertation Abstracts, NCJRS, PsychInfo, and Sociological Abstracts. Reviews of the effects of restorative justice on repeat offending and victims’ satisfaction with the handling of their cases were examined for references. Experts in the field were contacted.

SELECTION CRITERIA

The review includes only studies that employed a randomized design to test the effects of conferencing between at least one personal victim and one or more of their offenders on repeat offending or on victim impact, with the random assignment

following both offenders' and victims' consent to participate in an RJC if selected to do so. Ten eligible studies on three continents were identified, with a total of 1,879 offenders and 734 interviewed victims. The training for the RJC facilitators was provided by the same trainer in all ten trials, but that was not a criterion for selection. Cases were referred to the eligible experiments at various stages of the criminal justice process, including diversion from prosecution, post-conviction RCJs prior to sentencing, and post-sentencing RJC in prison and probation. The eligible tests included both violent and property crime, as well as youth and adult crime, with RJC offered as an alternative or as a supplement to prosecution in court. These variations provide a basis for moderator analyses as well as main effects on subsequent convictions (or in one case, arrests).

DATA COLLECTION AND ANALYSIS

The reviewers report the results of the ten eligible experiments identified. These experiments all reported post treatment data only of repeat crime measures at two years after random assignment (the only measurement period of offending common to the ten eligible trials). Measures for victim impact were also post-treatment, as measured by personal interviews with subsets of all victims who consented to random assignment.

All data analyses included in this review examined the effects of Intention-To-Treat (ITT), with wide variations in the percentage of both RJC and control cases receiving treatment as assigned. Many offenders assigned to prosecution, for example, failed to appear in court, just as many offenders assigned to an RJC failed to complete one. The analysis employs the ITT method to provide estimates of effectiveness under real-world conditions, at the expense of likely under-estimates of the efficacy of RJC when actually delivered. All studies reported effects on individual offenders and victims, while in all cases random assignment was done at the case level. In most trials the ratio of cases to offenders or victims was 1:1, while in others (the two Canberra experiments) that ratio ranged up to 1:1.25.

RESULTS

The evidence of a relationship between conferencing and subsequent convictions or arrests over two years post-random assignment is clear and compelling, with nine out of 10 results in the predicted direction and a standardized mean difference for the ten experiments combined (Cohen's $d = -.155$; $p = .001$). The impact of RJC on 2-year convictions was reported to be cost-effective in the 7 UK experiments, with up to 14 times as much benefit in costs of the crimes prevented (in London), and 8 times overall, as the cost of delivering RJC. The effect of conferencing on victims'

satisfaction with the handling of their cases is uniformly positive ($d = .327$; $p < .05$), as are several other measures of victim impact.

AUTHORS' CONCLUSIONS

RJCs delivered in the manner tested by the ten eligible tests in this review appear likely to reduce future detected crimes among the kinds of offenders who are willing to consent to RJCs, and whose victims are also willing to consent. The condition of consent is crucial not just to the research, but also to the aim of its generalizability. The operational basis of holding such conferences at all depends upon consent, since RJCs without consent are arguably unethical and breach accepted principles of restorative justice. The conclusions are appropriately limited to the kinds of cases in which RJCs would be ethical and appropriate. Among the kinds of cases in which both offenders and victims are willing to meet, RJCs seem likely to reduce future crime. Victims' satisfaction with the handling of their cases is consistently higher for victims assigned to RJCs than for victims whose cases were assigned to normal criminal justice processing.

Role of the Authors in Reviewed Studies

Two of the ten RCTs were designed, delivered and analyzed by research teams including three of the authors of this review (Strang, Sherman, and Woods). Independent authors gathered outcome data, analyzed and published results of the other eight trials, seven of which (Shapland et al, 2006, 2008) were operationally directed by two of the authors of this review (Strang and Sherman) and one of which (McGarrell and Hipple, 2007) was operated without contact with any authors of this review. One review author (Sherman) wrote the grant proposals and initial research designs for all ten eligible RCTs. None of the review authors had any conflict of interest in the results of the research, and three of the authors (Sherman, Strang and Woods) conducted the primary research for the only experiment out of ten included in this review that reported a backfiring effect of RJC's causing more crime.

1 Background

“Restorative justice” is a recent name for community practices that are thousands of years old (Braithwaite, 1998). The name refers to a broad range of practices, all of which define justice as an attempt to repair the harm a crime has caused rather than inflicting harm on an offender (Sherman and Strang, 2012). Other definitions emphasize a process of deliberation to decide what offenders should do that includes all people directly affected by a crime (Marshall, as quoted in Braithwaite, 2002: 11). Yet many procedures that lack such deliberation are also called restorative justice, including court-ordered community service, payments that offenders are required to make to their victims, and victim-offender mediation that excludes their families and friends. Recent programs in the UK have trained thousands of police to undertake “restorative disposals” or “community resolutions” that may involve negotiations on the street immediately after a crime has occurred, in which apologies are made and no further action is taken.

The diverse nature of these practices makes it difficult to answer the question of whether “restorative justice” defined so broadly works better than conventional justice, in either Common Law or Napoleonic legal traditions. The primary challenge, however, is empirical rather than conceptual. Most of the practices described as restorative justice have never been subjected to controlled field tests. Rigorous impact evaluations of restorative justice have been largely confined to a particular subset of programs, a subset we call “Restorative Justice Conferences” (RJC). This subset of restorative justice includes practices that have other names, including:

1. **“family group conferences,”** the traditional Maori practice which in 1989 became the primary basis for dealing with juvenile crime in New Zealand,
2. **“diversionary conferences,”** the name used in Australia to describe both juvenile and adult restorative justice as an alternative to prosecution, and
3. **“transformative justice,”** the name given to the approach by some trainers who use it to deal with conflict in employment and educational settings.

This subset is also similar to the Canadian practice of “sentencing circles,” which also builds on indigenous justice in a deliberation among those affected by crime, but which includes judges—unlike what we define as RJC.

1.1 DEFINITION OF RJC

Our definition of an RJC is this: a planned and scheduled face-to-face conference in which a trained facilitator “brings together offenders, their victims, and their respective kin and communities, in order to decide what the offender should do to repair the harm that a crime has caused” (Sherman and Strang, 2012: 216). This definition covers a homogenous group of programs inspired by the work of the Australian theorist John Braithwaite (1989) and the Australian trainer John McDonald, whose dialogue spread both the idea of RJCs and the opportunity for rigorous evaluations of them from Canberra to the US and UK from 1995 through 2005. Other training organizations have taught a similar method in English-speaking countries, emphasizing the following procedures to be followed by facilitators—most often police—trained to organize and convene an RJC that could last from 60 to 180 minutes or more:

- Facilitative discussion one-on-one with offenders and victims about what an RJC is, how it works, and whether they would consent to participate in one
- Scheduling of a conference at the victims’ convenience
- Seating all participants in a circle in a private space with a closed door, in settings ranging from police stations to prisons to community centers or schools
- Introducing all participants in terms of how they are emotionally connected to the crime under discussion
- Opening the discussion by asking offenders to describe the crime they committed
- Inviting victims and all participants to describe the harm the crime has caused
- When the harm has been fully described, inviting all participants, including the offender to suggest how the harm might be repaired, usually reaching a consensus on this question that is written up by the facilitator and signed by the offenders while all participants take a break for refreshments and informal conversation
- Filing the agreement with a court, a police unit, or some other institutional mechanism for encouraging compliance by the offender with the agreement.

This procedure has been used both in and out of criminal justice contexts, but all of the strong evidence of its effectiveness has been generated by comparisons to conventional criminal justice. These comparisons have been made with both juvenile and adult offenders who have accepted responsibility for their crimes in a wide range of offense categories, including burglary, serious assaults, vehicle theft, robbery and arson, at several points in the criminal process (Sherman and Strang, 2007, 2012):

- a) As post-arrest diversion from, and a substitute for, prosecution in court
- b) After a guilty plea in court, but before sentencing by a judge
- c) As part of a noncustodial sentence if requested by a probation officer
- d) After a period of imprisonment prior to release from prison

1.2 THEORETICAL BASIS

RJ Conferencing has strong theoretical connections to Braithwaite's theory of reintegrative shaming (1989), Tyler's theory of procedural justice (1990; Tyler and Huo, 2002), Sherman's theory of defiance (1993), Braithwaite's theory of responsive regulation (2002), and Collins' (2004) theory of interaction ritual chains. There is no causal theory that fully describes the manner in which conferencing might affect repeat offending and victims' satisfaction (see, e.g., Ahmed, et al, 2001).

Perhaps the closest theory to the predicted win-win effects of RJC's on offenders and victims is found in Collins (2004), whose theory is itself based partly on evaluations of RJC's. Using Durkheim's (1912) concept of "collective effervescence," Collins develops a causal model around the intense emotions of events like a RJC. Durkheim's concept denotes that the energy produced by a gathering of people changes their behavior in the aftermath of the gathering, as in a religious service that reaffirms a commitment to obey certain moral imperatives. Rossner (2011) provides some evidence that supports Collins' theory, but no tests have yet compared competing or complementary theories of why RJC's can affect offending behavior and victim outcomes.

Collins' theory also provides the basis for limiting the present review to crimes in which an identifiable person has been harmed as a victim. RJC's have been tested on both the "victimless" crime of driving with blood alcohol levels over prescribed limits, and on the crime of shoplifting against corporate victims (Sherman and Strang, 2012). In neither test did the offender confront anyone with whose suffering they could empathize, suffering which the offender had personally caused. While we have reported the results of these tests elsewhere (Sherman and Strang 2012), we exclude them from the present review on the theoretical grounds that they do not

share the fundamental bio-psychological conditions of an RJC with cases in which a harmed person faces an offender (Sherman and Strang, 2011). This decision has no effect on the conclusions reached below (since the two excluded studies reach opposite conclusions with each other about RJC effects), but it does set a theoretically sound basis for the future addition of new studies to updates of this review. The best interpretation of the available evidence to date on RJC is that the evidence offers an assessment of a policy rather than a theory. This conclusion is especially warranted by the wide range of delivered treatments in the wake of random assignment. In medical terms, the available evidence includes virtually no *efficacy* trials, under controlled conditions, guaranteeing high levels of delivery of the program elements described above. Rather, the available evidence reports what are best described as *effectiveness* trials under real-world conditions. Future research that creates greater consistency of delivery of RJC elements may yield different, and possibly stronger, effect sizes than those reported in this review.

2 Objectives

The objectives of this review are to answer two primary questions:

- a) What is the effect on repeat offending of a policy of attempting RJs with consenting victims and offenders?
- b) What are the effects of a policy of attempting RJs with consenting victims and offenders on various measures of whether victims have been restored to their circumstances prior to the crime?

Because frequency of criminal convictions (or arrests) is a crude indicator of the amount of harm caused by crime, the review also sought information indicating the *seriousness* or cost of crime as a measure of impact on repeat offending.

3 Methodology

3.1 CRITERIA FOR CONSIDERING STUDIES FOR THIS REVIEW

This review of the effects of RJC's was limited to studies that had all eight of the following characteristics:

- 1) Study was reported in the English language.
- 2) Study tested a Restorative Justice Conference (RJC) as defined above.
- 3) Study used random or quasi-random assignment to the RJC condition and a control condition of criminal cases in which an arrest or other official action had been imposed.
- 4) Study involved offender samples that committed crimes against one or more identifiable individuals.
- 5) Study involved offenders and victims in the study had consented to accept random assignment to either participating in an RJC or doing without one, prior to random assignment. Study provided data on the frequency of post-random assignment criminal convictions of offenders or re-arrest for two years after random assignment.
- 6) Study reported data that enabled the calculation of an intention-to-treat (ITT) effect, rather than treatment as delivered effect.
- 7) Study was conducted after 1994.

These criteria are justified below. As Braithwaite (1998, 2002) suggests, the restorative justice label embraces a wide range of similar programs that have very different dynamics. These differences could create heterogeneity in the program content that would limit the face validity of our systematic review. A leading example is Victim-Offender Mediation (VOM) programs. VOM is more structured than conferencing, and mediators play a much more prominent (and more negotiator-like) role in controlling the discussion in VOM than conference facilitators play in RJC's. While supporters are sometimes involved, VOM may consist only of the victim, the offender, and the mediator. In VOM, the mediator negotiates between the two parties; the victim and the offender may never meet face

to face. The primary focus of VOM is often material restitution rather than emotional restoration or reconciliation (Umbreit et al, 1994). For similar reasons, the eligibility criteria for this review excludes Victim-Offender Reconciliation Programs (VORP) (Peachey, 1989) and ‘circle sentencing,’ in which a judge talks to stakeholders about the appropriate penalty for a crime before formally imposing a sentence (Stuart 1996).

Random assignment generally provides the best means for eliminating selection bias, as well as other rival hypotheses, in assessing the effects of a policy (Cook and Campbell, 1979). Non-random comparison groups are abundant in restorative justice evaluations (McCold, 1998; Miers, et al, 2001), but are arguably plagued by biased selection of cases that were deemed more “appropriate” for RJC than cases to which they were compared—either historical or matched controls, including some studies in which those who refused RJC were compared to those who agreed.

The requirement for identifiable victims is justified by the very different dynamics observed in RJC with and without a victim present. Qualitative evidence indicates far lower levels of emotional intensity and offender remorse in cases without personal victims than in cases where personal victims are engaged (see also quantitative observational data in Strang, et al, 1999). In terms of interaction ritual chain theory (Collins, 2004), the level of collective effervescence in the conference appears far lower in RJC without a personal victim: conference length appears much shorter, tears appear less often. Victimless conferences may also be less traumatizing for the offender than the description provided by Peter Woolf (2007), a high-frequency burglar who suffered nightmares and racing thoughts for years after a long RJC where two of his victims vehemently expressed their anger.

The issue of consent prior to random assignment shapes a decision made to exclude two experiments conducted in Bethlehem, Pennsylvania (McCold and Wachtel, 1998), in which over half of the cases randomly assigned to RJC failed to comply with the treatment as assigned. The high refusal rate followed the use of a procedure in which consent was sought after random assignment rather than before. This decision not only adversely affected the internal validity of the test. It also affected the external validity of the test to cases in which participants agree to attend an RJC. Because random assignment preceded the agreement, the population randomly assigned did not match the target population to which the study could be generalized. This review is limited to studies that define the target population as an eligibility criterion prior to random assignment.

The decision to use *frequency* of subsequent recidivism as the outcome for offenders is driven by both policy and pragmatism. The policy issue is whether a measure of

prevalence of future offending is a reliable indicator of public benefit without taking frequency into account. Since total harm to the public corresponds more closely to the number of crimes committed than to the number of active criminals committing those crimes, the review chooses the former. It thus provides a clearer guide to policy by preferring frequency counts over the “one or more crimes” measure of proportion of offenders re-offending.

As a matter of pragmatism, frequency of convictions is also a more statistically powerful and less confusing way to measure impact in small samples. It thereby reduces bias due to low power, and the potential confusion that underpowered tests may cause to policymakers. Shapland et al’s (2008: 27) meta-analysis of the seven UK experiments in RJC, for example, shows consistent benefits of restorative justice using both prevalence and frequency measures, both of which have similar effect sizes. Yet because of its lower power levels, the prevalence analysis fails to achieve statistical significance in meta-analysis. Shapland et al’s (2008: 27, Figure 2.6) frequency analysis, in contrast, shows significance levels well within conventional thresholds ($p = .013$), again with the same effect sizes as in the prevalence analysis. Yet the authors have repeatedly encountered confusion among UK policymakers about the meaning of prevalence vs. frequency, and a reluctance to make policy based on “mixed” results. This review chooses to clarify the findings by use of the single measure (Piantadosi, 1997: 128) that the authors recommended from the outset of the first trials of RJC: frequency of offending (see Sherman et al, 2000).

The preference for convictions where available is also pragmatic, since 7 of the 10 experiments eligible for this review reported on no other measure of repeat offending. Only one of ten experiments (McGarrell and Hipple, 2007) reported no data on convictions, using arrests as the only repeat offending measure. Given the juvenile status of the offenders in that one exception, this may be a distinction without a difference as data on juvenile arrests in Indiana appear to be recorded on a similar basis as juvenile convictions are reported in the UK data. A similar pragmatic criterion limited the outcomes to post-treatment differences only, which is all that was reported for 8 of the 10 eligible experiments.

The two-year window of outcome assessment for offending effects is selected in accord with the recommendations of the Coalition for Evidence-Based Policy, the National Research Council, and the Office of Management and Budget, all in the United States.

Finally, the use of an intention-to-treat (ITT) criterion is, in the authors’ view, essential for this review. It is only by using ITT that we can meet our objective of testing a policy of attempting RJCs, not just the effects of completing RJCs. Given

the costs inherent in each attempt, it is far more policy-relevant to the public interest to understand the overall benefit of attempting to deliver RJs in relation to the total cost of the attempts—including both successes and failures.

3.2 SEARCH STRATEGY FOR IDENTIFICATION OF STUDIES

The authors searched reference lists, contacted other authors, conducted electronic searches, and examined all reports related to restorative justice in the program of the American Society of Criminology from 1997 to 2012. Published and unpublished studies were considered. While some databases were restricted to particular periods of time, electronic searches were not otherwise limited by date. Indexes were searched in which non-English publications were expected to appear, but only reports written in English were considered for the review.

In 2012, one author electronically searched 15 databases related to criminal justice, law, and related areas of social science. The most common search was applied to databases indexed by Cambridge Scientific Abstracts; these databases were searched using the following terms: ((restorative AND (justice OR sentenc*)) OR (mediate OR mediation OR restitution OR conferencing) AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend* OR recidiv* OR victim) AND (ab=random* OR ab=control*). All databases searched and the particular terms used to search each database are listed in Table 1.

Table 1: Electronic searches

Database	Search(es)	Hits
Bibliography of Nordic Criminology (BNC)	("restorative justice" or mediation or conference or restitution) AND (criminal OR offender OR perpetrator) AND (random or randomly or randomized)	63
Criminal Justice Abstracts	((restorative AND (justice OR sentenc*)) OR (mediate OR mediation OR restitution OR conferencing AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend* OR recidiv* OR victim) AND (ab=random* OR ab=controll*)	20
Criminal Justice in Denmark (CJD)	("restorative justice" or mediation or conference or restitution) AND (criminal OR offender OR perpetrator) AND (random or randomly or randomized)	7
Dissertation Abstracts	((restorative AND (justice OR sentenc?)) OR (mediate OR mediation OR restitution OR conferencing AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend? OR recidiv? OR victim) AND ab(random? OR controll?)	106
IBSS: International Bibliography of the Social Sciences	((restorative AND (justice OR sentenc*)) OR (mediate OR mediation OR restitution OR conferencing AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend* OR recidiv* OR victim) AND (ab=random* OR ab=controll*)	5
Index to Foreign Legal Periodicals	1) kw criminal OR kw offender OR kw perpetrator; 2) kw restorative justice or kw mediation or kw conferencing or kw restitution; 3) #1 & #2	5
NCJRS Abstracts Database (NCJRS Virtual Library)	((restorative AND (justice OR sentenc*)) OR (mediate OR mediation OR restitution OR conferencing AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend* OR recidiv* OR victim) AND (random* OR controll*)	154
PAIS International	((restorative AND (justice OR sentenc*)) OR (mediate OR mediation OR restitution OR conferencing AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend* OR recidiv* OR victim) AND (ab=random* OR ab=controll*)	38
PILOTS Database	((restorative AND (justice OR sentenc*)) OR (mediate OR mediation OR restitution OR conferencing AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend* OR recidiv* OR victim) AND (random* OR controll*)	45
Political Research Online	Subject: restorative justice	14

PsychInfo	((restorative AND (justice OR sentenc*)) OR (mediate OR mediation OR restitution OR conferencing AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend* OR recidiv* OR victim) AND (ab=random* OR ab=controll*)	53
Social Services Abstracts	((restorative AND (justice OR sentenc*)) OR (mediate OR mediation OR restitution OR conferencing AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend* OR recidiv* OR victim) AND (ab=random* OR ab=controll*)	2
Social Work Abstracts	((restorative AND (justice OR sentenc*)) OR (mediate OR mediation OR restitution OR conferencing AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend* OR recidiv* OR victim) AND (ab=random* OR ab=controll*)	3
Sociological Abstracts	((restorative AND (justice OR sentenc*)) OR (mediate OR mediation OR restitution OR conferencing AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend* OR recidiv* OR victim) AND (ab=random* OR ab=controll*)	4
Worldwide Political Science Abstracts	((restorative AND (justice OR sentenc*)) OR (mediate OR mediation OR restitution OR conferencing AND ((criminal OR offender OR perpetrator) AND victim))) AND (reoffend* OR recidiv* OR victim) AND (ab=random* OR ab=controll*)	1
Total Hits		519

3.3 SELECTION OF STUDIES

One author checked titles and abstracts to identify studies that could be excluded based on information provided in the title or abstract. When a study could not be excluded based on that information, more information was obtained by retrieving the article or by contacting the authors.

The search identified articles in languages other than English. The authors are not aware of any completed or ongoing RCTs that have not been reported in English, but the authors are unable to conclude that none would be identified by combing these articles or by conducting a broader search.

Two authors extracted information from the full text of articles when published reports were available. Other information was obtained directly from investigators, including the authors and their colleagues in the primary studies.

Given the decision to limit eligible studies to RCTs, which were small in number, the studies were not compiled in a coded format.

3.4 DATA MANAGEMENT AND EXTRACTION

Data on repeat offending and on victim impact were extracted from each of the completed studies. Where this information was missing from the published reports, the reviewers requested it directly from the original investigators.

3.5 OUTCOME MEASURES

The authors would have preferred the use of before treatment-after treatment frequency analysis as the most logically sound test of intervention effects on recidivism. Pragmatically, however, only two studies offered before-and-after frequency analysis, while ten of them offered only post-treatment frequency measures. To examine outcomes from the maximum number of experiments, the authors decided to employ the “highest common denominator” allowing comparative analyses of effect sizes: two-year post-treatment differences in the frequency of criminal convictions per offender for nine of the studies, and of arrest in Indianapolis.

3.6 EFFECT SIZE ESTIMATES AND MODERATOR ANALYSES

The reviewers used Comprehensive Meta-Analysis v.2 (Borenstein et al, 2005) to analyze frequency of conviction with the standardized mean difference (Cohen’s *d*). Outcomes were meta-analyzed using traditional inverse-variance weighted meta-analysis. In all cases, a random effects model was assumed *a priori*. The *Q*-test was used to measure for heterogeneity across effect sizes.

Samples of criminal cases may vary on many dimensions, each of which poses a challenge in a systematic review that integrates the findings of diverse tests. Examining the effects of RJs across a wide range of offenses and offender types is not unlike examining the effects of aspirin across a wide range of diseases, including cancer, heart disease, influenza, sunburn, and syphilis. Further, the character of RJ conferences may change in relation to the populations and problems studied. There is no *a priori* reason to expect any intervention to be equally or consistently effective across all conditions, particularly when the intervention is an interaction among people rather than a drug. The reviewers attempt to avoid generalizations about

included studies that would mislead readers about the effects of conferencing under tightly defined specific conditions.

Studies of conferencing vary in several ways, including offender age, offense type, location in the criminal justice process, type of comparison interventions, measures of dependent variables, period of follow-up, and percentage of cases in which the intervention is delivered as assigned. Some of these differences may also be related. With a small universe of eligible studies, the best we can do is to present moderator analyses in a variety of ways.

DRAFT

4 Results

4.1 DESCRIPTION OF STUDIES

In all, 15 RCTs and one study that appeared to be an RCT were considered for the review. Six were excluded and ten remained (See Appendix A for rationale for exclusions). The eligible studies we included covered five jurisdictions on three continents, across a range of decision points in the criminal justice system, with a total of 734 interviewed victims and 1,879 offenders accepting responsibility for their crimes. The main characteristics of each experiment are described in Table 2.

Table 2: Case and Offender Characteristics of Experiments Included in the Review, By Experiment

Location of Experiment	Time Period	Evaluators	Offense type(s)	Point in Justice System	Control	% of RJC delivered as assigned	N
1. Canberra	1995-2000	Sherman and Strang	Violence, under 30	Diversion from Prosecution	Prosecution	79%	121
2. Canberra	1995-2000	Sherman and Strang	Property, under 18	Diversion from Prosecution	Prosecution	68%	249
3. Indianapolis	1997 ¹	McGarrell and Hipple	Violence & Property under 14	Diversion from Prosecution	Other Diversion Programs; VOM	80%	782
4. London	2001-5	Shapland et al	Robbery Over 18	Post-plea, presentence	No RJC presentence	85%	106

¹ McGarrell and Hipple do not report the date on which they stopped random assignment.

5. London	2001-5	Shapland et al	Burglary Over 18	Post-plea, presentence	No RJC presentence	85%	186
6. Thames Valley	2001-5	Shapland et al	Assault over 18	In prison, prerelease	No RJC in prison	73%	103
7. Thames Valley	2001-5	Shapland et al	Assault over 18	On probation	No RJC on probation	87%	63
8. Northumbria	2001-5	Shapland et al	Violence & Property under 18	Diversion to a police caution with RJC	Diversion to caution with no RJC	92%	165
9. Northumbria	2001-5	Shapland et al	Property over 18	Post-plea, presentence	No RJC presentence	77%	105 ²
10. Northumbria	2001-5	Shapland et al	Violence Over 18	Post-plea, presentence	No RJC presentence	(77%)	
Total							1880

4.2 ASSESSMENT OF METHODOLOGICAL QUALITY OF INCLUDED STUDIES

4.2.1 Randomization

None of the included studies reported problems with randomization.

Randomization was in the hands of the research staff in the Canberra RISE (Reintegrative Shaming) Experiments (nos. 1, 2 in Table 2) and in the seven UK experiments (nos. 4-10). Those nine experiments had RJC facilitators calling a remote research office for random assignment after identifying details of eligible cases were recorded by the research team. In contrast, in the Indianapolis experiment (no. 3), randomization was the responsibility of the operational partner, the Juvenile Court.

The Indianapolis experiment and the UK experiments randomized *offenders* to interventions. In the Canberra because some crimes involve multiple offenders, the experiments randomized *cases*; however, data are reported for individual offenders and victims, not cases. This approach violates the principle of “analyse as you randomize,” but the data are not available at the level of case averages or central

² Shapland et al (2008: 25) combined the two Northumbria Magistrates’ Court experiments in reporting the rate of RJC delivery as assigned, hence the same data are reported for each.

tendencies. This was not a serious issue because the ratio between the case and the individual in these two studies was only 1:1.25.

4.2.2 Attrition from treatment as assigned

As Table 2 indicates, none of the trials delivered the interventions exactly as intended. In some cases, offenders failed to appear in court. Some conferences were not held because offenders failed to cooperate. In some cases conference facilitators failed to organize a conference.

In the Canberra Youth Violence Experiment (#1), 85% of offenders were treated as their cases were assigned; 49 of 62 offenders (79%) assigned to conferencing received conferencing and 54 of 59 offenders (92%) assigned to court went to court.

In the Canberra Juvenile Personal Property Experiment (#2), 76% of offenders were treated as their cases were assigned; 83 of 122 offenders (68%) assigned to conferencing received conferencing and 105 of 127 offenders (83%) assigned to court went to court.

The Indianapolis Experiment (#3) with juvenile first offenders yielded an 80% completion rate for RJC-assigned cases (322 of 400) and a 61% completion rate (233 of 382 cases) for the control group programs of diversion from prosecution (McGarrell and Hipple, 2007: 230).

In the seven UK trials, analysis was reported on the basis of “invitation to treat” (Shapland et al 2008: 12, FN 23). The completion rates of conferences (but not of control group cases) was reported by Shapland et al (2006: 25) to vary between 100 percent for the London robbery experiment and 83 percent for the Thames Valley prison experiment.

When examining recidivism, offenders assigned to conferencing were analyzed as if they attended conferences, even if they were eventually dealt with in the same way as the control group, or not at all. While it limits the ability of this review to describe the effects of conferencing on recidivism for those subjects who attended conferences, this method of analysis (“intention-to-treat” - ITT) is not biased by any differential attrition (Piantadosi, 1997: 276-78). Despite any remaining debate over whether an ITT is preferable to a treatment-on-treated approach, the ITT approach is consistent with the objective of the review. The ITT approach measures the likely effects of introducing a *policy* of conferencing in which not everyone assigned to conferencing would complete the RJC. Given the high rate of attrition in all of the included studies, the authors concluded that “per protocol” analysis, or an analysis of “treatment-on-the-treated,” would bias the review.

With one exception, Table 2 shows that the experiments had at least 70% of the offenders assigned to RJC actually participate in them. With virtually no crossover of control groups receiving RJC, there is a reasonably logical basis for expecting different outcomes from the two randomly assigned groups. The single exception (#2) in meeting the threshold, in a way provides even more assurance for that point: it is the only experiment in ten in which assignment to RJC was followed by less than 73% delivery of RJC. With only 68% of RJC-assigned offenders getting RJC, one could speculate that the result was due to inadequate dosage of the treatment. A more plausible explanation, however, may be that a large number of Aboriginals were referred into that experiment, and for them the effect of RJC was extremely toxic: an over 200% increase in before-after differences in repeat offending (Sherman, et al, 2006).

More important may be the relatively small range in which RJC was delivered as assigned. Table 2 shows that seven out of ten experiments had between 77% and 87% of the RJC-assigned cases treated-as-assigned. As the basis for an effectiveness estimate to be generalized to real-world conditions, the narrow range suggests that most RJC programs may deliver at similar rates and with similar effects, assuming a similar mix of referred cases and similar cultural backgrounds.

4.2.3 Time at risk

In most of the ten experiments, imprisonment was rarely used in either the RJC or control group cases (though in the case of #6, the offender was already in prison). The two exceptions to this rule were the London robbery and burglary experiments. In these two studies, the offenders had extensive criminal records of prior convictions and instant convictions for serious crimes, so some time in prison for both experimental (RJC-assigned) and control offenders was often mandatory under sentencing guidelines. The procedure employed by Shapland and her colleagues (2008) was to eliminate randomly assigned cases from the analysis if the offenders had served the entire two years after random assignment in prison. Since there were no significant differences in the likelihood of a prison sentence for most of the time period of random assignment, this analytic decision was not likely to create a bias *between* treatment groups. What it did create, however, is a highly heterogeneous mix of days at risk *within* each treatment group. By including a case if there was even one day of liberty in the community, or $365 \times 2 = 730$, a very wide range of risk periods was allowed, without standardizing the rate of convictions per days at risk by dividing the numerator of convictions by the exact number of days at liberty. The rate of repeat offending per day at risk was therefore highly variable, even among offenders with one reconviction, yet the two-year frequency is presented almost as if it is equivalent by days at risk. Since there is no way for a secondary reviewer to create a standardized measure, the only choice is between inclusion or exclusion of these findings from eligibility for the analysis.

The inclusion of these two studies in the meta-analysis reduces the estimates of effect size relative to excluding them, as we report below under sensitivity analysis. It is therefore a more conservative procedure to retain them in describing the main effects of the meta-analysis than to remove them.

Other issues of method could be addressed, but not improved upon, in a secondary analysis. Given what is known about these ten experiments, they would appear to provide a reasonably homogeneous basis for data synthesis.

4.3 META-ANALYSIS OF REPEAT OFFENDING

The primary criterion of the effect of RJC on crime is the frequency of repeat offending over the two years after random assignment. In the meta-analyses presented below, the post-treatment measure of repeat offending is criminal convictions in all tests except Indianapolis, for which the measure is repeat arrests. We first calculated the odds ratios (OR) for the outcomes and then converted these OR into standardized differences of means (d) using the logit method.

The Key for the studies identified by three letters in the forest plots is listed below, with the number corresponding to the chronological list of the experiments in Table 2, arranged here by their effect size in reducing crime in Figure 1:

JPP	=	Juvenile Property Crime, Canberra, Australia,	No. 2
LOR	=	London Robbery (street crime), UK,	No. 4
LOB	=	London Burglary, UK,	No. 5
TVP	=	Thames Valley Prison, UK assault cases,	No. 6
IND	=	Indianapolis juvenile crime, USA,	No. 3
NCP	=	Northumbria Court Property crime, UK,	No. 9
TVC	=	Thames Valley Community sentence, UK, assaults	No. 7
NFW	=	Northumbria Final Warning for juveniles, UK	No. 8
JVC	=	Juvenile Violent Crime, Canberra, Australia,	No. 1
NCA	=	Northumbria Court Assault,	No. 10

Figure 1 below shows that the average effect of RJC is to reduce crime. More precisely, across 1,879 offenders in all 10 eligible experiments, the average effect size is .155 standard deviations less repeat offending among the offenders in cases randomly assigned to RJC than among the offenders in cases assigned not to have an RJC. The 95% confidence interval for this effect lies between only .06 standard deviations less crime and .25 standard deviations less crime. This means that the

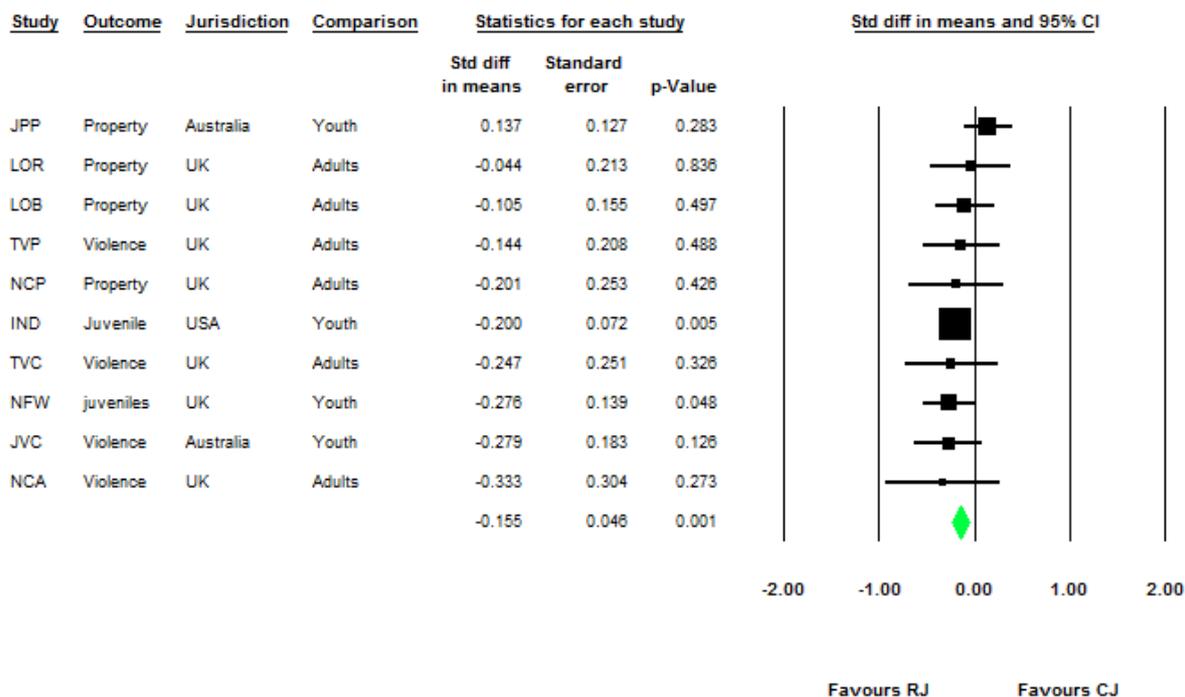
average effect across all these experiments is highly unlikely to be a chance finding ($d = .155, p = .001$).

Put another way, only one out of the ten experiments shows a statistically significant effect,--but 9 out of 10 of the experiments show less crime with RJC's than without them. Either of those calculations alone could be misleading. But when the average effect size across all ten studies is calculated—including one in which there was *more* crime with RJC's (but not significantly more)—the *pattern* of findings can be described as statistically “significant” and favoring the benefits of RJC's. That means, in this case, that there is only a one in a thousand chance that the pattern in Figure 1 could have occurred by chance.

What is difficult to convey about these findings is how many crimes were prevented, or how big the effect of RJC's is likely to be in practical terms. The percentage differences associated with the ten experiments range from 7% to 45% fewer repeat convictions or arrests. This may help practitioners to grasp how much crime that would mean with the kind of offenders they might consider using RJC's with. But an even better way to judge the practical value of these differences is to use the cost-of-crime prevented data presented in section 5 below.

Figure 1 - Effects of RJC on Frequency of Repeat Offending, 2-year Follow-Up Period

COMBINED EFFECTS SIZES FOR STUDY OUTCOMES



Meta-Analysis Random Effects Model, $Q = 7.754$, $df = 9$, $p < 0.559$

4.4 MODERATOR ANALYSES

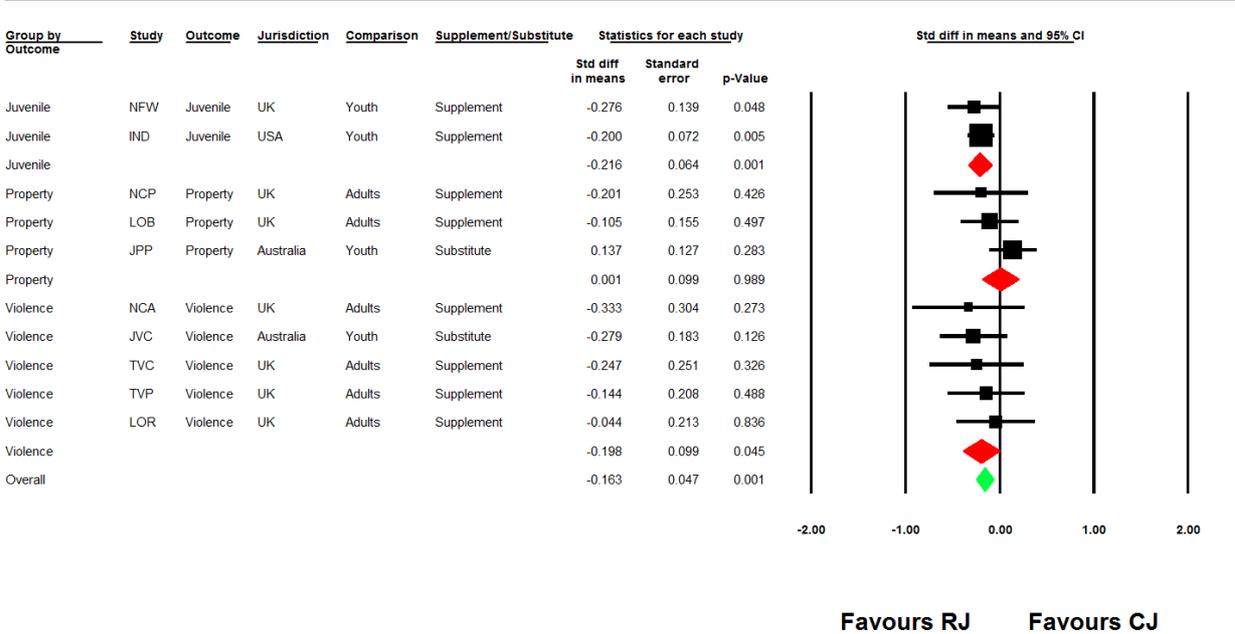
The overall meta-analysis of the ten experiments can be unpacked to learn whether RJC's work better with some kinds of samples, or in some kinds of experiments, than others. These different ways of sorting the experiments are called “moderator analyses,” because they can reveal whether some third factor is “moderating” or changing the findings. By “third factors” in this review we mean the age of the offenders, or the kinds of offenses they were arrested for. That could suggest, for example, that if RJC's were used with only the kinds of cases associated with a third factor, it would get much better or worse results than the average effects across all ten experiments. Because the ten experiments vary widely in the third factors they represent, it is important to probe whether the overall average is being driven up or down by one or more of those factors. That is the purpose of presenting Figures 2-3 below.

Half of the experiments in the sample, for example, tested RJC's with violent crimes. Figure 2 shows what the average effect of RJC's is on just violent crimes. (Two others had a mix of violent and property crimes: Indianapolis and Northumbria Final

Warning). The average effect of RJC for experiments limited to violent crimes was .2 standard deviations. That is an effect size that is 28% larger than the effect of RJC for all ten experiments. This means that, on average, RJC appear to work better for violent crimes than for all crime types in these ten experiments combined, but because that difference is not statistically significant ($Q = 1.021, P = .9$) it must be treated with caution. Three of the other five experiments used samples of property crimes only. Figure 2 shows that RJC have far less effect, on average, in these property crime experiments than in the violent crime experiments. The average effect appears to be very close to zero. This result could have been different with a different set of property crimes or offenders, and it is hard to generalize on the basis of just three experiments. Nonetheless, there seems to be something very different about the impact of RJC for property crime than for violent crime.

Many public officials say that RJC are more appropriate for juvenile offenders than for adults. Yet the findings from this Review suggest otherwise, at least for offenses with personal victims. In Figure 2 we see that the average effect of RJC in six experiments with all adults is .150 standard deviations. Yet as shown in the Figure, we see that the average effect of RJC on experiments with juveniles is only .119. The difference in effect size between adult and juvenile offenders is not large. But it is nonetheless in the opposite direction from the conventional wisdom.

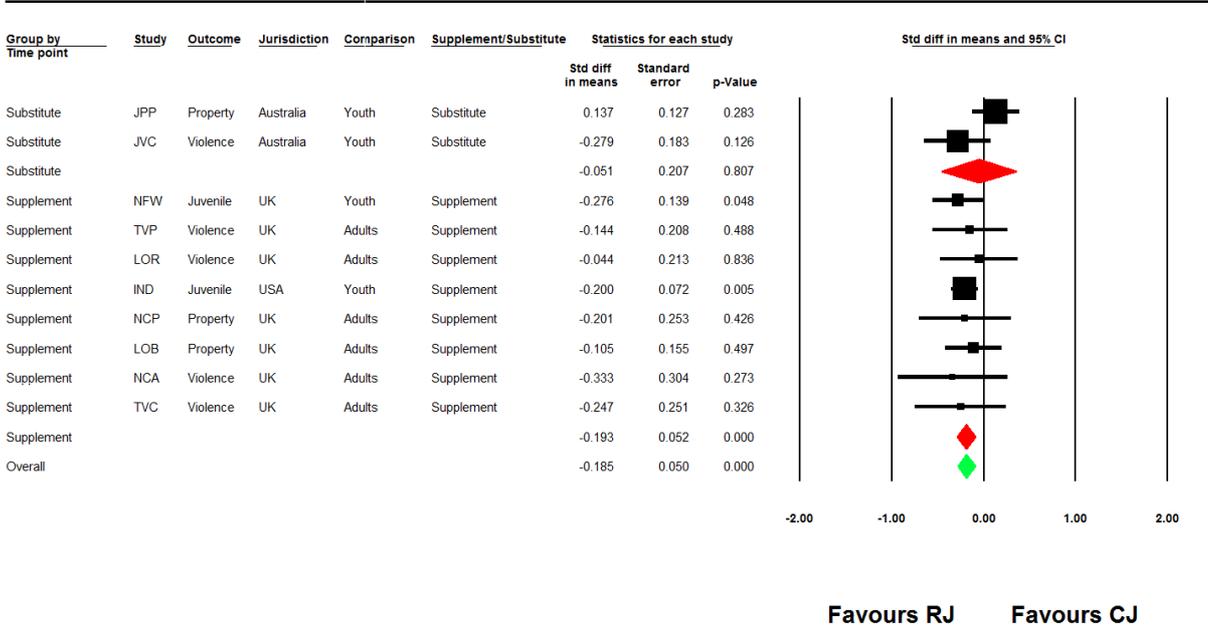
Figure 2: Crime Type as Moderator of Study Outcomes



Juveniles $Q = 0.233, df = 1, p < 0.630$; Property $Q = 2.244, df = 2, p < 0.326$; Violence $Q = 1.021, df = 4, p < 0.907$; Between Group $Q = 3.574, df = 2, p < 0.167$

One of the major policy debates in restorative justice is whether it should merely supplement conventional justice (CJ), or replace it altogether. In Figure 3 we see that the average effect of RJC is larger in the 8 experiments when it is used as a supplement to conventional justice than for the average effect for all ten experiments (.19 vs. .15), but this difference is not statistically significant ($Q = 0.447$, $P = 0.50$). Thus while the average effect for using it as a substitute may appear to be lower, the broad range of effect sizes in the two tests of RJC as a substitute leaves us too uncertain about its average effect. Put another way, both the worst and second-best results in the entire sample are found in the substitutional category. How much lower the effect of using RJC as a diversion from conventional justice can, somewhat, be shown from the two studies. The moderator analysis in Figure 3 shows that on average, the two experiments in Canberra with personal victims had almost no effect (.001 standard deviations difference) on the frequency of repeat offending. It also shows that the individual studies went in opposite directions, canceling each other out. Moreover, the effect of the diversion of violent crimes to RJC was .279 standard deviations, one of the largest benefits in the entire meta-analysis. Based on these two studies alone, there may still be potential for using RJC as a diversion rather than as a supplement. More research will be needed for a reliable comparison of substitutional and supplemental uses of RJC.

Figure 3: Effects of RJC as Supplement or Substitute to Conventional Justice on Frequency of Repeat Offending, 2-year Follow-Up Period



RJC as Substitute $Q = 3.491$, $df = 1$, $p < 0.062$; Supplement $Q = 1.483$; $df = 7$, $p < 0.983$; Between Group $Q = 0.447$; $df = 1$, $p < 0.504$

4.5 SENSITIVITY ANALYSES

In this section, we report a series of tests for whether the results presented above are “sensitive” to the inclusion or exclusion of certain kinds of tests, which may reflect certain kinds of biases that could in turn limit the generalizability of the results. The points we examine are the effects of the authors as evaluators, the effects of using arrests (in Indianapolis) in a meta-analysis that uses convictions in all nine other experiments, and the effect of excluding from the sample offenders who had no time at risk to re-offend because they were in prison for the entire follow-up period of two years after random assignment (or treatment).

4.6 EVALUATOR EFFECTS.

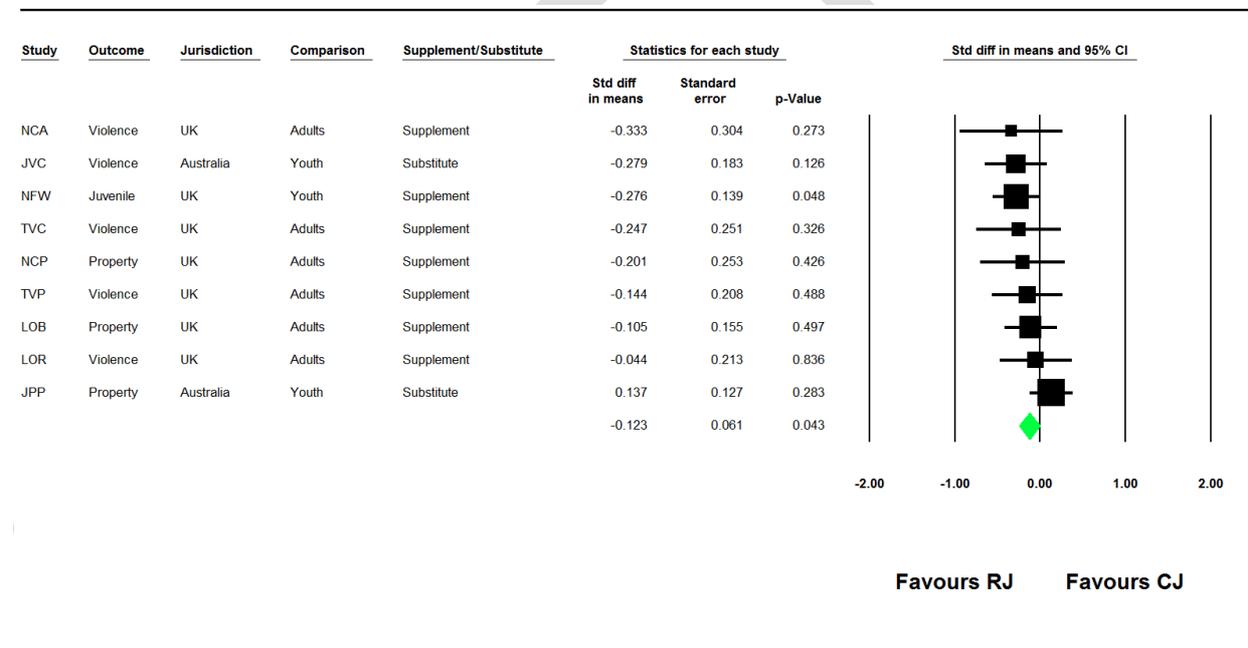
Some readers may wonder whether the inclusion of studies in a meta-analysis in which the primary research was done by the analyst has an impact on the conclusions. The answer in this study is yes, but not in the expected direction. Petrosino and Soydan (2005) and Eisner (2009) have both suggested that there is an effect in which evaluations associated with people who develop programs are likely to show better outcomes than evaluations in which no developer is a collaborator. The definition of a “developer” may be somewhat problematic, and the authors do not think of themselves as RJ developers. Trainers like John McDonald seem more appropriate for that title. Yet “developer” of the RCTs is how Sherman and Strang were described by the UK government in the UK experiments that were independently evaluated by Joanna Shapland and her team of evaluators.

It is difficult, but not impossible, to examine that issue within this review. It is true that at least one of the authors had some association, however distant, with all ten of the experiments. But there is one bright line to examine. In only two of the experiments did the authors of this review gather the outcome data and perform the analysis that produced the results analyzed above. In all eight of the other experiments, that task was done by independent analysts. As it happens, the difference between the two is exactly the same as the difference between the two experiments using RJC as a substitute (developed and evaluated by Sherman and Strang) and the eight experiments with evaluators independent of Strang and Sherman as developers. And as Figure 3 shows, the eight experiments with independent evaluators reported *better* results for RJC effects on repeat offending than the experiments in which review authors also did the analysis. If there is a bias created by inclusion of the review author’s own evaluations, it is a bias against showing RJC to be effective.

4.6.1 Arrests vs. Convictions.

Figure 4 addresses the question of whether the results of this review are sensitive to the use of arrests in one experiment, while the others report convictions. It displays the effect of 9 experiments, omitting the Indianapolis study—which accounted for over one-third of all the offenders in the review. The effect of removing Indianapolis is to reduce the effect size of RJC somewhat, but not to change the direction or the statistical significance of the result. Compared to the effect for all ten experiments (.15), the effect size of .12 without Indianapolis is close enough to conclude that the result is not sensitive to any aspect of including or excluding this study from the meta-analysis.

Figure 4: Effects of RJC on the Frequency of Criminal Convictions, 2-year Post-treatment Follow-up Period



4.6.2 Time at Risk.

As noted above at 4.2.3, two of the ten studies - the London burglary and robbery pre-sentence experiments - used a procedure that included all offenders who were out of prison for any period of time during the two years after date of random assignment, from one day to two years minus one day, without controlling for variation in time at risk (Shapland, et al, 2008). They did, however, have reported effect sizes based on the evaluators' decision to delete any cases in which the offender was incarcerated for the entire two-year followup period. We elected to include these studies because the result of doing so was apparently to reduce the overall mean effect size of the ten available tests. Because we could not make any secondary attempts to standardize repeat conviction (or arrest) rates by days at risk

(out of prison) within the two-year followup, the only choice was between inclusion or exclusion of these two London experiments evaluated by Shapland et al (2008). Figure 5 shows that the mean effect size of RJC on repeat offending when these two London studies are removed, so that all studies consistently have no deletions for any reduced level of time at risk. The standardized mean difference across only the eight studies was $D = .165$, or slightly higher than the mean effect for all ten studies (see Figure 1). This difference was due to a lower effect size of adding RJ to criminal sentencing in the two London experiments than in the two Thames Valley experiments, which were confined to assault cases but also had very serious injuries. Figure 6 shows that the mean effect size for the two studies that deleted randomly assigned cases in which offenders spent the entire two-year followup period in custodial punishment was only $.08$, or far lower than the overall mean. This does not indicate that the results of RJC for robbery and burglary cases are necessarily less effective. It could simply mean that more time is needed to examine the impact of RJC in such serious cases. More years of followup could provide more time for offenders to re-offend (or not), potentially even showing bigger effects on the cost of crime than found in experiments with less serious instant offenses. The point is that we simply cannot tell what the long-term effects would be without further followup.

Lest it appear that the smaller effect sizes may be due to less time in prison in the RJC group than in the conventional justice group, we can cite a separate study conducted by Strang, Barnes, Sherman, Bennett and Inkpen (2005), which found no significant differences between the RJC and conventional justice groups in the London experiments in either the prevalence of sentences to time in prison or the mean number of days sentenced. The study was conducted because of an initially higher rate of prison sentences for the RJC group than for the cases randomly assigned to the conventional justice (no-RJC) group. This difference flattened out by the end of the enrolment of all the cases the program randomly assigned. While not all of those cases were included in the Shapland, et al (2008) evaluation, the vast majority were. If there is any difference, it would be more prison time with RJC than without it. Prison cannot therefore explain why the effect of RJC would be lower in these experiments, as opposed to being higher due to a “boost” from more incapacitation from imprisonment.

Figure 5. Effects of RJC on the Frequency of Repeat Offending (Without Deletions for Time at Risk), 2-year Post-treatment Follow-up Period

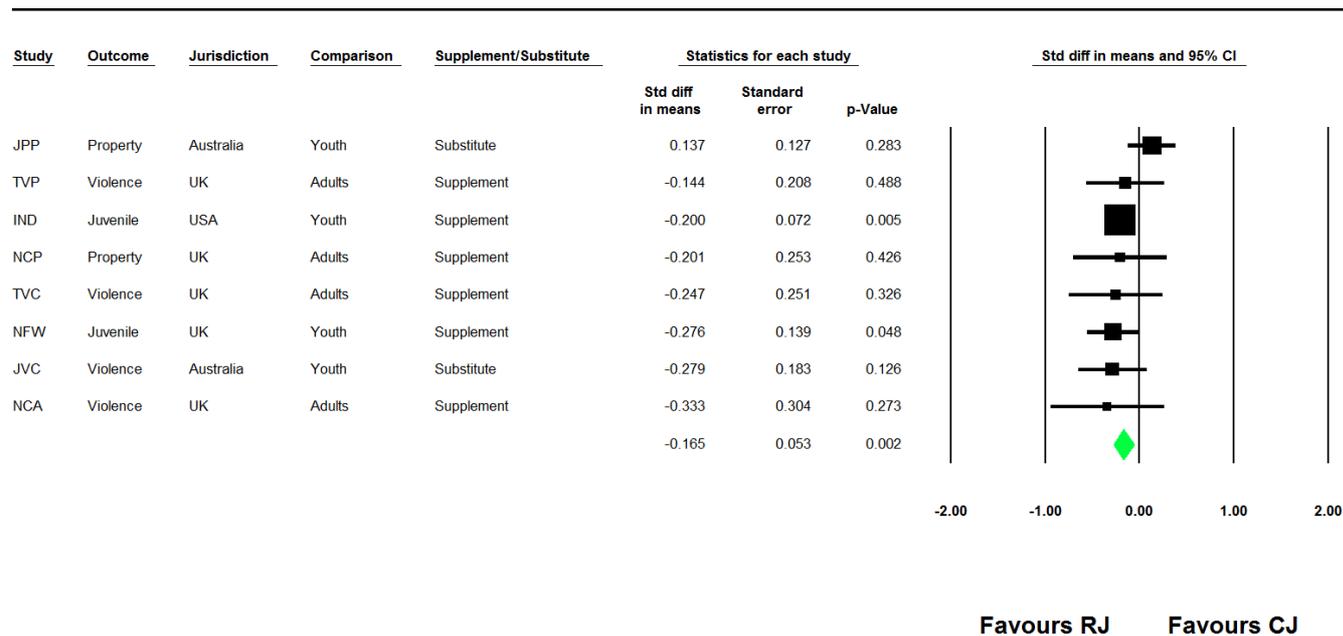
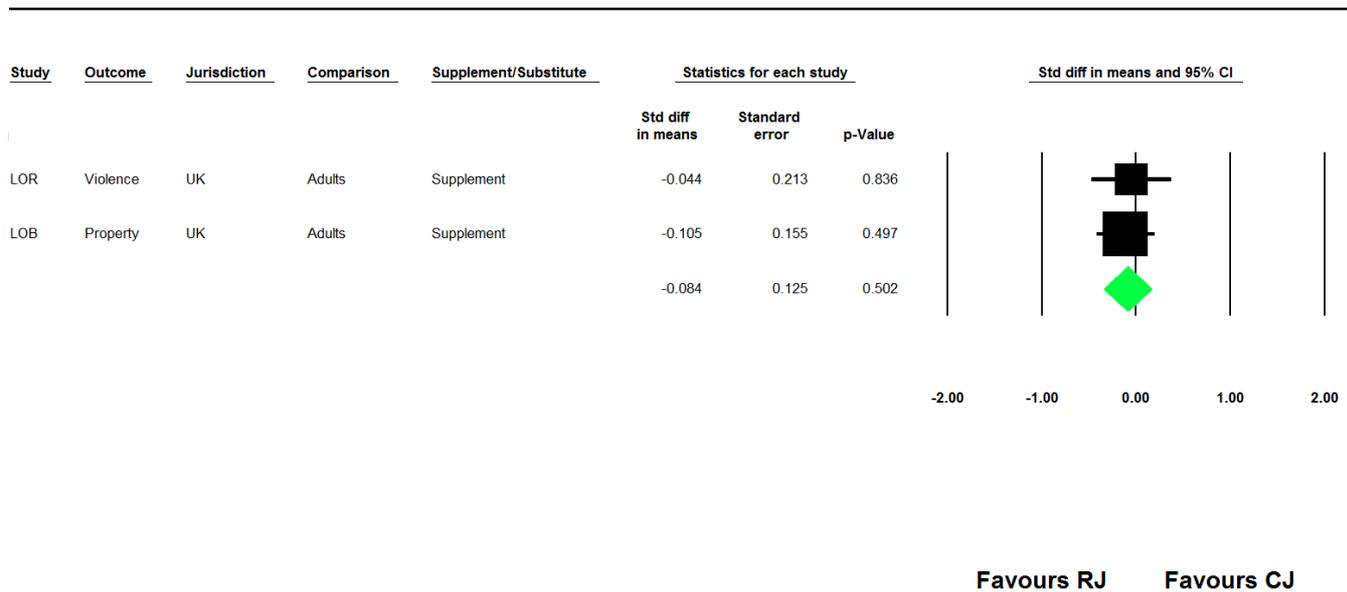


Figure 6: Effects of RJC on the Frequency of Repeat Offending (With Deletions for No Time at Risk), 2-year Post-treatment Follow-up Period



4.7 VICTIM IMPACT MEASURES: NARRATIVE FINDINGS FROM THE TEN INCLUDED STUDIES

Although RJC is often described as a victim-centered approach to justice, evaluations of effects and effectiveness often are not reported as clearly or extensively for victims as for offenders. This deficit exists both for methodological reporting – numbers of victims in the study, number of interviews, response rates – and for reporting of outcome measures. These problems arise because all the studies reported in this review – the two Canberra (Reintegrative Shaming) experiments, the seven United Kingdom (UK) experiments conducted by the Justice Research Consortium (UK), and the Indianapolis study of young offenders – use cases or offenders as their unit of random assignment and analysis: victims are ‘attached’ to the cases randomly assigned and the random assignment sequence does not create comparability across victims in the same way that it does for offenders.

In all these studies, primary attention was focused on offender effects, especially on re-offending, rather than victim effects. Also, while criminal justice records must identify each offender arrested in a case, there is no imperative to record each victim affected: indeed, there will never be any official record of victims who do not come to the attention of the police or where the identification of complainants/victims in a case is highly arbitrary, e.g. a spouse or children in a burglary. In addition, in some cases where the eligibility criteria are met in respect of the offense and the offender, there may be no identifiable victim, e.g. possession of a concealed weapon. Thus, an objective of including in the sampling frame for any study all victims associated with the randomly assigned cases (or offenders) is not practically possible and the research designs of the included studies can only be quasi-experimental with respect to victim effects.

4.7.1 Intention-to-treat and disappointment

All data analyses of victims included in this review examined the effects of ITT). This allows the inclusion of RJC-assigned victims who never experienced RJC even though they were told that their cases would be dealt with this way. These victims were often disappointed when their expectations were not met; indeed, they turned out to be some of the most dissatisfied of all. Including their views on the basis of their assignment rather than their experience allows us to understand the likely views of victims expecting but failing to receive RJC in the event of a policy to make it available universally.

4.7.2 Victim response rates

One way of assessing the possible extent of bias in what is essentially a convenience sample of victims associated with each case is to estimate response rates on the most

relevant denominator and numerator. The appropriate denominator is limited to those cases for which at least one victim has been identified; the appropriate numerator can be either the number of cases for which one or more victims have been interviewed, or the number of individual interviews achieved. In the Canberra experiments in juvenile property crime and youth violent crime, both victim response rates have been reported (Strang 2002, p 77-78) but in the UK experiments (Shapland et al 2007) reported only the former. In the Indianapolis study (McGarrell 2001) only the number of victim interviews has been reported and they are a sub-sample of all victims involved in the study as the interviews could be conducted only towards the end of the study period.

Table 3 summarised the best available information from all the studies. In the interests of comparability, Canberra data are reported on the same basis as the data published on the UK experiments. It should be noted, however, that these data are not available for individual UK experiments: Shapland et al (2007) reported victim responses by site, and each site had two or more experiments.

Table 3: Response rates for victim interviews

Site	N of Cases with 1+ Contactable Victims	N of Interviews (Individuals)	Response Rate % (Cases)
London:			
RJC group	119	76	59%
Control group	125	54	42%
Northumbria:			
RJC group	146	104	69%
Control group	120	79	64%
Thames Valley			
RJC group	59	36	58%
Control group	72	33	44%
Canberra Juv Property			
RJC group	66	71	88%

Control group	72	80	92%
Canberra Violence	Youth		
RJC group	45	45	89%
Control group	38	36	84%
<hr/>			
Indianapolis			
RJC group	n.a.	42	n.a.
Control group	n.a.	50	n.a.
<hr/>			

4.7.3 Victim outcome measures

Measures of victim effects varied in type and in detail across the ten experiments. The only measure consistently reported goes to the question of victim satisfaction at the conclusion of the disposition, whether by RJC or the control treatment. This is sometimes disaggregated into various aspects of satisfaction and sometimes reported as a global measure.

Beyond ‘victim satisfaction’ broadly defined, only the Canberra experiments report on specific aspects of material and emotional restoration, though many of these questions were asked only of the RJC victims and are not included in this review because there are no measures for the control group. The same is true in the UK experiments. It should also be noted that findings for the UK experiments sometimes are reported as aggregate victim measures by the site of the study and sometimes reported across all seven experiments (i.e. a composite measure for all RJC-assigned compared with all control-assigned). This means that there are no offense-specific victim data in the UK experiments that compare victims assigned to the experimental condition with victims assigned to the control condition. In the Indianapolis study (McGarrell et al 2000, McGarrell 2001, McGarrell & Hipple 2007), reporting of victims’ views is limited to a few dimensions of satisfaction, several with no actual numbers or percentages reported, and with no response rates reported.

This review is therefore able to identify a limited number of dimensions of victim impact on which responses by victims assigned to the RJC and control groups can be compared in at least two experiments: material restoration, emotional restoration,

satisfaction with the process, dissatisfaction with the process, desire for revenge, and post-traumatic stress symptoms.

4.7.4 Material restoration

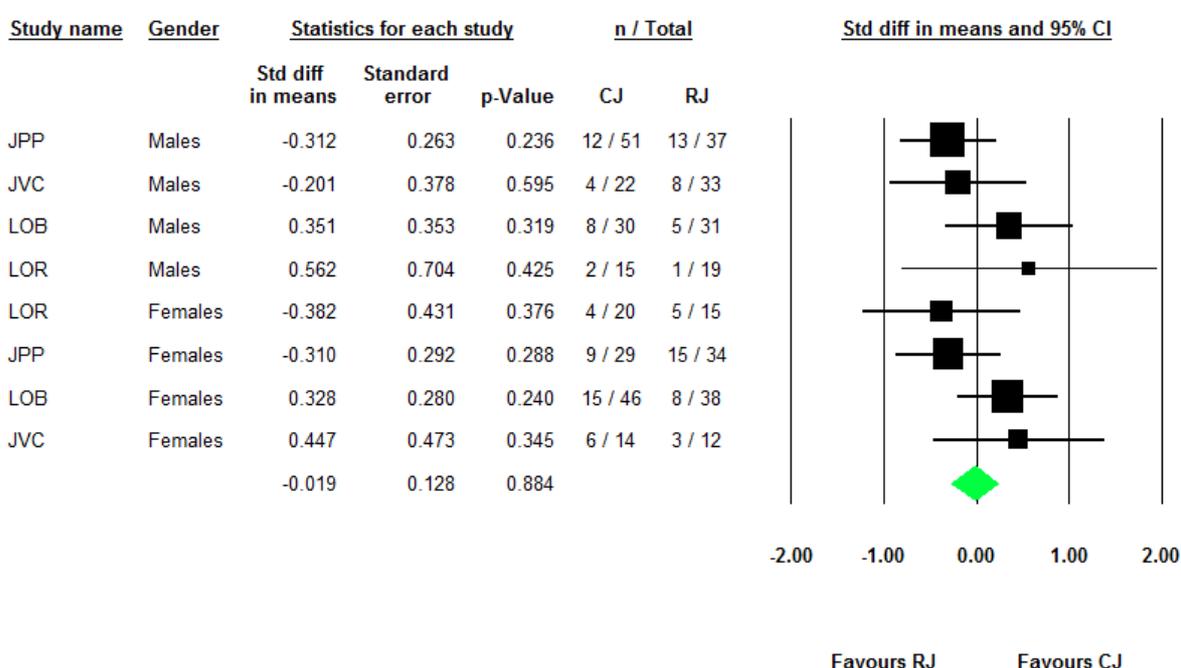
Data only are available for this measure in the two Canberra experiments. However, although material restoration is a legitimate and significant part of a restorative process, victims in the two Canberra experiments indicated they did not always regard it as being of primary importance: this was fortunate as so few were awarded financial restitution in either RJC or court. Of the 47 percent of court-assigned victims who wanted money, 12 percent received it; of the 38 percent of RJC-assigned victims who wanted money, 16 percent received it. The RJC victims, however, had the possibility of receiving other forms of restitution in their outcome agreement: 11 percent of them accepted work or other benefits offered by their offenders in lieu of money.

4.7.5 Emotional restoration

The emotional harm victims suffer from crime can take many forms. One of them concerns self-blame for the crime's occurrence. Victims in the two Canberra experiments and the UK burglary and robbery experiments were asked whether they blamed themselves for what had happened. Figure 7 shows that there were no consistent differences in self-blame between the treatment groups or across locations (reprinted from Sherman, et al, 2005).

Figure 7: Victim Self-Blame

VICTIM SELF-BLAME



$Q=7.469; df=7, p<0.382$

An important measure of emotional restoration for victims is a sense of safety or, conversely, a fear of victimisation. In Canberra, 18 percent of court-assigned victims anticipated that their offender would repeat the offense on them, compared with five percent of RJC-assigned victims ($p<0.005, d=.78$). Among *property victims*, three times as many court as RJC victims believed the offender would do so (21 percent vs. 7 percent, $p<0.05, d=.70$); among *violence victims*, more than five times as many court as RJC victims believed the offender would do so (11 percent vs. 2 percent, $p=0.01, d=.99$).

When Canberra victims were asked whether they anticipated their offender would repeat the offense on another victim, a significantly higher percentage of court than RJC victims believed they would do so (55 percent vs. 35 percent, $p<0.005, d=.47$). Among *property victims*, a significantly higher percentage of court than RJC victims believed this would be the case (54 percent vs. 31 percent, $p<0.05, d=.53$). Similarly, among *violence victims*, significantly more court victims than RJC victims believed this would be the case (58 percent vs. 40 percent, $p<0.01, d=.402$).

Perhaps the most significant factor in emotional restoration relates to whether victims receive an apology from their offender, and how they rate the sincerity of the apology offered. Almost 90 percent of all Canberra victims said that they wanted an

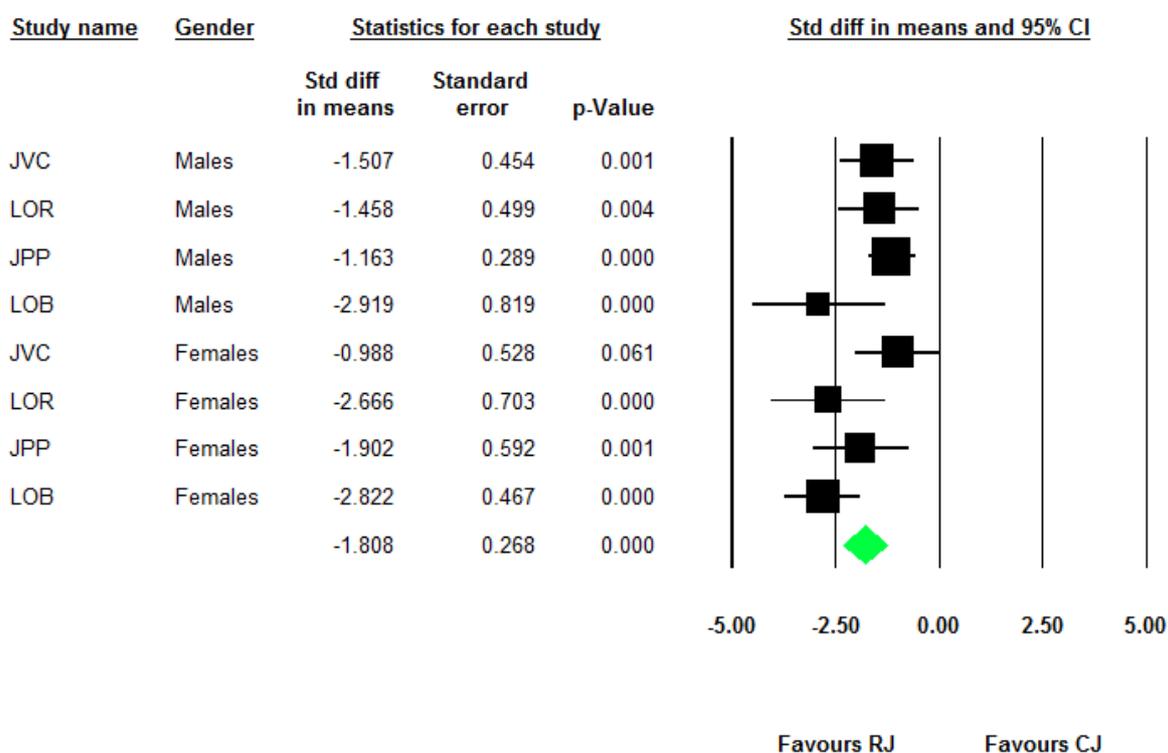
apology but there were great differences between the court-assigned victims and the RJC-assigned victims when they were asked whether they had received one. Of the victims whose cases were assigned to RJC, 72 percent said they had received an apology (and 86 percent of those who had actually attended an RJC), compared with 19 percent of the court victims ($p < 0.000$, $d = 1.33$). (In none of the court cases was the apology part of the court outcome but rather negotiated separately, whereas it was almost always part of the RJC outcome).

When victims in the UK burglary and robbery experiments were asked about apologies, again there were great differences between the RJC-assigned and the court only-assigned. In burglary, 96 percent of the RJC victims received an apology compared with 7 percent of the court victims ($d = 3.18$). In robbery, all of the RJC victims received an apology compared with 14 percent of the court-only victims. Figure 8 displays these data in a meta-analysis reprinted from Sherman et al 2005.

There was also a significant difference between the groups when they were asked to rate the sincerity of the apology: in Canberra, 58 percent of the RJC *violence victims* believed it was 'sincere' or 'somewhat sincere', compared with 11 percent of the court victims ($d = 1.33$); 55 percent of the RJC *property victims* said that it was 'sincere' or 'somewhat sincere' compared with 10 percent of the court victims ($d = 1.32$). In the UK experiments, 79 percent of the RJC *robbery victims* rated it as sincere compared with 11 percent of the court-only victims ($d = 1.88$); 57 percent of the RJC *burglary victims* said it was sincere compared with 7 percent of the court-only victims ($d = 1.58$).

Figure 8: Victim Received a Sincere Apology:

VICTIM RECEIVED A SINCERE APOLOGY

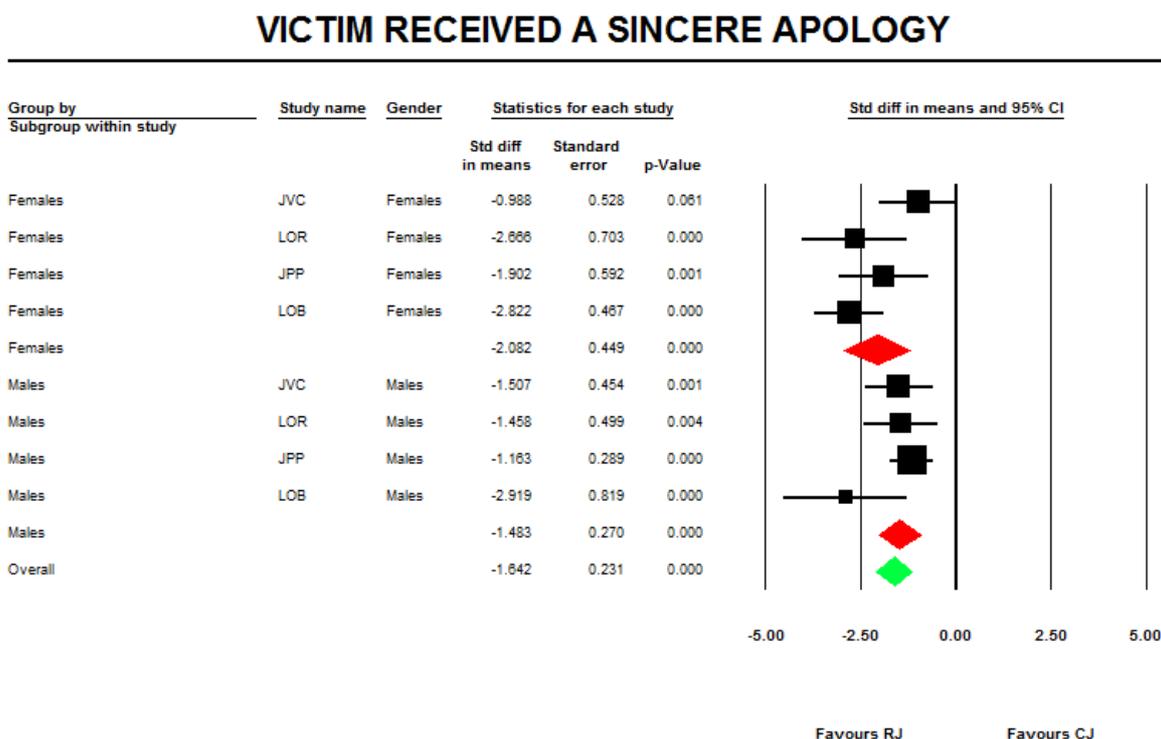


$Q=15.606; df=7, p<.029$

These findings confirm that courts often neglect the non-material dimensions of victimisation, while RJC is moderately successful in delivering the emotional restoration victims seek, and especially in providing a forum for the transaction of apologies.

Furthermore, given the significant heterogeneity, we conducted a moderator analysis for this outcome based on gender. As shown below in Figure 8, the magnitude of the difference for female victims was much larger than for male victims ($d=-2.082$ compared to $d=-1.642$, respectively).

Figure 9: Victim Received a Sincere Apology: Gender analysis



Females $Q = 7.588$; $df=3$, $p<0.055$; Males $Q=4.175$; $df=3$; $p<0.243$

4.7.6 Satisfaction with the process

All ten experiments compare RJC-assigned and control-assigned victims on measures of satisfaction, though they do so in ways not easily comparable. They will therefore be reported separately.

4.7.6.1 Canberra Experiments

All victims in the property and violence experiments were asked whether they were satisfied with the way their case was dealt with by the justice system: 46 percent of the court-assigned victims vs. 60 percent of the RJC-assigned were satisfied ($p<0.05$, $d=.327$). (Significantly more of those who actually experienced RJC were satisfied, compared with those whose cases were dealt with in court: 70 percent vs. 42 percent, $p<0.001$). There was virtually no difference between the responses of property and violent victims here: for *property victims*, 61 percent of the RJC-assigned and 46 percent of the court-assigned were satisfied ($d=.34$ and for *violence victims*, 60 percent of the RJC-assigned and 44 percent of the court-assigned were satisfied ($d=.36$).

All victims were also asked whether they were pleased that their cases were dealt with in the way they were (whether by RJC or by court), rather than the alternative

disposition. Significantly more RJC-assigned victims than court-assigned victims agreed they were pleased with their treatment (69 percent vs 48 percent, $p > 0.005$, $d = .472$). When *violence victims* were asked, 66 percent of RJC-assigned and 58 percent of court-assigned were pleased ($d = .188$). For *property victims*, 70 percent of RJC-assigned and 43 percent of court-assigned were pleased ($p > 0.005$, $d = .623$).

4.7.6.2 UK Experiments

Victims in the seven UK experiments were asked how satisfied they were with what the criminal justice system did about their offense. No data are available for the experiments separately but in the aggregate 72 percent of RJC victims said they were satisfied compared with 60 percent of the control victims ($d = .30$). In addition, significantly more of the RJC victims than the control victims said their treatment made them feel more secure (Chi-square = 8.926, $df = 1$, $p = 0.003$) (Shapland et al 2011, p147).

All UK victims also were asked whether they felt the criminal justice process was fair. Aggregated across all seven experiments, 73 percent of RJC victims felt it was fair compared with 61 percent of the control victims ($d = .30$).

Finally, all UK victims were asked whether, as a result of their treatment, their view of the criminal justice system had changed (no base rates available). Again, aggregated across all experiments, 34 percent of RJC victims said it was more positive compared with 28 percent of the control victims ($d = .16$).

4.7.6.3 Indianapolis

In the Indianapolis Juvenile Restorative Justice Experiment (McGarrell et al, 2000), 92 percent of RJC-assigned victims of juvenile property and violent offenders reported that they were satisfied with the way their case was handled compared with 68 percent of control-assigned ($d = .93$).

4.7.7 Dissatisfaction with the process

Dissatisfaction with the treatment victims received turns out to be as important an outcome measure as the indicators of satisfaction. Only Canberra victims, however, were asked to assess their negative feelings about their treatment.

When they were asked whether the way their case was dealt with made them feel angry, 14 percent of the Canberra RJC-assigned *property victims* agreed, compared with 29 percent of the court-assigned ($d = -.51$), and 24 percent of the RJC-assigned *violence victims* compared with 39 percent of the court-assigned ($d = -.39$). When they were asked whether they felt bitter about the way they were treated, 9 percent of the RJC-assigned *property victims* compared with 13 percent of the court-assigned and said they felt bitter ($d = -.23$), and 22 percent of the RJC-assigned

violence victims compared with 31 percent of the court-assigned ($d=-.26$). On the global satisfaction/dissatisfaction measure, 21 percent of the court-assigned victims vs. 20 percent of the RJC-assigned were *dissatisfied* ($p<0.05$, $d=.33$).

4.7.8 Desire for revenge

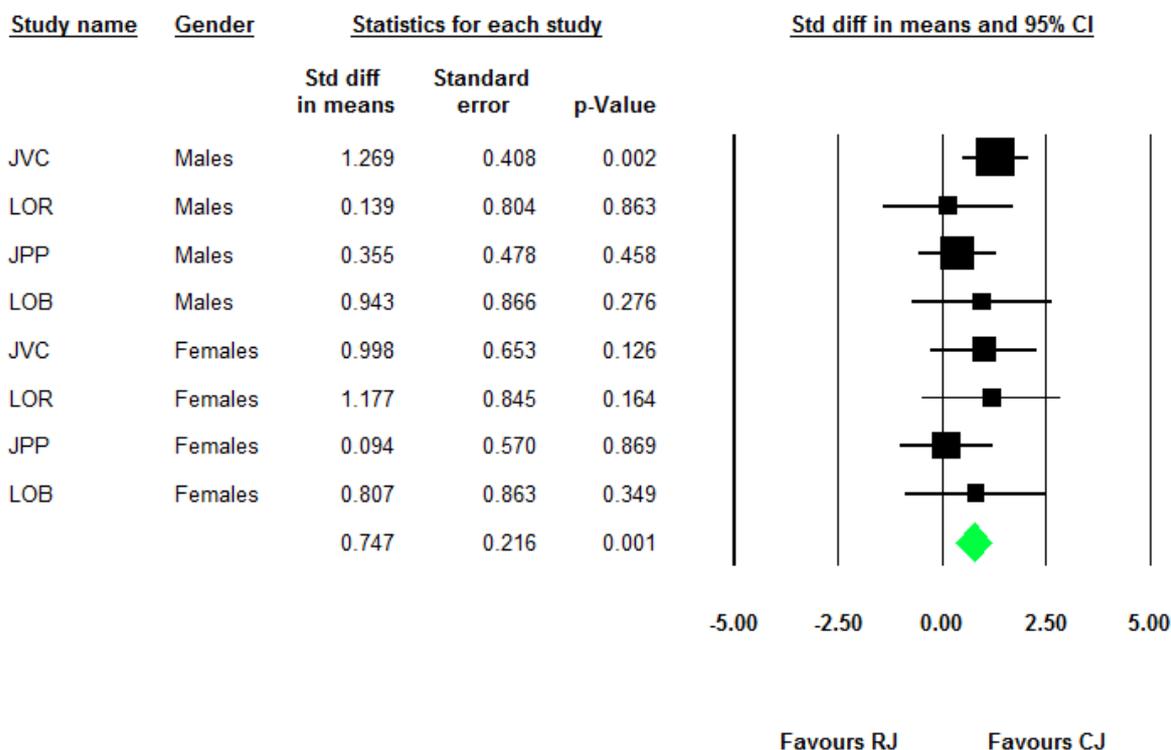
An underestimated aspect of victimisation is the personal anger victims sometimes feel towards their offenders which, especially in the case of violent crime, may be translated into a desire to physically harm them. This aspect was explored in the two Canberra experiments and in two of the UK experiments, London robbery and London burglary (Figure 10).

In the Canberra experiments victims were asked whether they would harm their offenders themselves if they had the chance. Only a small percentage of *property victims* said they would do so (9 percent of the court-assigned and 6 percent of the RJC-assigned, $d=-.24$), whereas with *violence victims*, 45 percent of the court – assigned said they would do so, compared with only 9 percent of the RJC-assigned ($d=-1.17$).

In the UK experiments, the prevalence of desire for revenge was also lower among the RJC-assigned victims than the control-assigned victims in the London burglary and robbery experiments (Angel 2005, Sherman et al 2005). None of the *burglary victims* assigned to RJC said they wanted to harm their offenders compared with 5 percent of the control group victims; only 3 percent of the *robbery victims* assigned to RJC wanted to harm their offenders compared with 14 percent of the control group ($d=-.92$). These findings strongly suggest that RJC can succeed in assuaging the feelings of vengeance felt by many victims of violent crime towards their assailants. Figure 10 displays these effects in a reprint from Sherman et al (2005).

Figure 10: Desire for Revenge

DESIRE FOR REVENGE



$Q=4.663; df=7, p<0.701$

4.7.9 Post-traumatic stress symptoms

Post-traumatic stress disorder is a clinical condition describing pathological reactions to events causing psychological trauma. Such an event is defined as one in which the victim experiences an event as one likely to cause serious injury or death and reacts with emotions of fear, helplessness and horror (American Psychiatric Association, 1994). It can result in many adverse consequences including reduced quality of life, work impairment and other sequelae.

Some victims may suffer post-traumatic reactions that fall short of a diagnosis of PTSD but nonetheless may suffer components of the diagnosis which constitute post-traumatic stress symptoms (PTSS). These symptoms have been measured in a sub-set of the victims in the UK robbery and burglary experiments using a standardised clinical test, the Impact of Events (Revised) Scale (IES(R)) (Weiss & Marmar, 1997). Victims in these experiments were interviewed twice, the first time soon after the disposition of the case and the second time six months later. At the initial interview, RJC-assigned victims, where RJC was in addition to normal court

proceedings, were found to have reduced PTSS compared with the control-assigned victims, whose cases were dealt with by court alone ($p=.07$, $d=.308$). This remained the case at the follow-up interview ($p=.07$, $d=.341$) (Angel 2005). Although there was no baseline assessment of victims' psychological health prior to randomisation, it appears likely RJC has a beneficial outcome for victims experiencing PTSS.

4.8 COST EFFECTIVENESS

The measurement of harm caused by crime to the community is generally under-developed. This review has relied primarily on the inadequate measure of frequency of crimes, in which all crimes are counted equally. In this framework, a murder is equal to an auto theft; a rape is equal to a burglary. Treating crimes of such disparate weight with equal seriousness is, on reflection, offensive to fundamental human values. We do not sentence people to prison for equal terms for these unequal crimes. Neither can we be content with evaluating the impact of crime as if all crimes caused equal harm.

In seven of the experiments included in this review, the evaluators (Shapland, et al 2008) took the highly original and important step of giving widely varying weight to each crime for which offenders were convicted in the two-year follow-up period. They did this in two ways, both of which had been developed by the UK government. The first method was to use a scale of crime severity. The problem with that method is that as adopted by the Home Office at the time, the scale was truncated at 10 to 1. That is, the maximum difference between murder and any other crime was limited to 10 times greater seriousness for murder than, for example, a pickpocket taking a wallet with £5 in it. Such a "flat" scale communicates the differences in crime seriousness no better than saying that a \$1,000,000,000 annual salary is only ten times greater than a salary of \$30,000.

The second and far more accurate method that Shapland and her colleagues (2008) used was the Home Office calculations of the cost of crimes, based on empirical research for average crime costs over samples of many of the most common kinds of crime. This method, developed by DuBourg and his colleagues (2005), employs a range of tens of thousands of pounds or dollars between the lowest and highest cost crimes. As Shapland et al. (2008) applied it to the data in the RJC experiments they evaluated, it created a far more sensitive metric for the evaluation of RJC effects on offenders. Evaluating impact in this way produced much larger effect sizes, greater statistical power, and differences in effect sizes from the measure based on frequency of crimes counted equally.

The most striking evidence of how the impact assessment can be changed substantially by using costs rather than counts is found in their results for London. While the experiments with robbery and burglary offenders in London yielded small and non-significant effect sizes of RJC on the frequency of reconvictions, the cost-effectiveness ratios in London were the highest of any UK experiment in RJC. As compared to the 61% lower frequency of violent offenders given RJC in the Northumbria Magistrates' court (Sherman and Strang, 2012: 231), the London robbery experiment had only 8% fewer reconvictions and the London burglary experiment had only 16% fewer reconvictions for RJC cases than controls. But when the cost of crime prevented in *London* was compared to the running costs of delivering RJC (excluding startup costs of a new project), the ratio was £14 in the cost of crime prevented for every £1 spent on delivering RJC. In *Northumbria*, across all 3 experiments (as reported by Shapland et al 2008), the cost-benefit ratio was only £1.2 in cost of crime saved for every £1 invested in police delivering RJC. Yet all of the cost calculations in the UK sites were statistically significant, even where they were not for comparing counts of crimes. In *Thames Valley*, the benefit across the two experiments aggregated was reported at £2 in costs of crime prevented for every £1 spent on RJC.

It is worth noting that in their analysis of costs and benefits, Shapland (et al) 2008 made two key distinctions. One was between running costs and startup costs; the other was between total costs and costs only to the criminal justice system. The difference between ongoing, year-in-year out “running” and the one-time, initial “startup” costs is an important issue for external validity. Startup costs may vary much more widely than running costs, especially in terms of working out the inter-agency arrangements needed to establish a process of recruitment of cases and delivery of treatment. Startup can take a year or more, with the costs depending on how many people are assigned to the job of implementing a very different way of processing criminal cases. It is arguably more appropriate to focus on the running costs, which indicate what can be the costs after a startup period—no matter how costly or low-cost the startup may be. The labor costs for delivery are much lower than for the construction of the process, and of greater interest to those who would like to run restorative justice as a long-term strategy.

The second distinction the Shapland et al (2008) cost-benefit analysis draws is between total costs versus criminal justice system costs. We highlight here the total costs, since health and welfare costs are often born by taxpayers and personal costs to victims are of concern to the public interest. Some officials, however, prefer a closed system of cost analysis, in which the focus is on how much money a criminal justice reform can save for the criminal justice budget. For those who prefer that approach, they may find the data in Shapland et al (2008), which clearly show less

benefit (to criminal justice alone) in return for RJC costs than for the total estimated costs of crime.

DRAFT

5 Authors' conclusions

RJCs delivered in the manner tested by the ten eligible tests in this experiment appear likely to reduce the future frequency of detected and prosecutable crimes among the kinds of offenders who are willing to consent to RJCs, when victims are also willing to give consent to the process. The condition of mutual consent is crucial not just to the research, but also to the aim of its generalizability. The operational basis of holding such conferences at all depends upon consent, since RJCs without consent are arguably unethical. The Review's conclusions are appropriately limited to the kinds of cases in which RJCs would be ethical and appropriate. Among the kinds of cases in which both offenders and victims are willing to meet, RJCs seem likely to reduce frequency and (with less data) costs of future crime. Victims' satisfaction with the handling of their cases is consistently higher for victims assigned to RJCs than for victims whose cases were assigned to normal criminal justice processing.

5.1 IMPLICATIONS FOR PRACTICE

The effects of RJCs on the frequency of repeat offending are especially clear as a supplement to conventional justice, with less certainty about its effects when used as a substitute. Yet RJCs may be seen as most appealing when they can both reduce crime and save money—starting with diversion from expensive court processes. The use of restorative processes in this way has grown rapidly in some countries without rigorous testing, sometimes by citing the evidence from using RJCs as a supplement. Cost-saving goals have apparently strengthened the appeal of RJ in theory, but without the kind of evidence reviewed here.

Readers should be well-advised that nothing in the present review provides any evidence in support of immediate “community resolutions” or “restorative resolutions” using restorative principles. That does not mean that the review shows such a quick-fix approach cannot work. It simply means that the time-consuming preparations for a two to three hour conference led by a specialist cannot be compared to a brief interaction at the scene of an incident or shortly thereafter, often with minimal victim involvement. The present review shows only the effects of

formal RJ conferences arranged well in advance so that all persons affected by a crime may have a chance to attend.

When RJ conferences are conducted as they were in the experiments included in this review, there can be a high confidence of good results with violent crime, and somewhat less confidence with property crime. The evidence suggests that with serious offenders with long criminal records, the delivery of RJC also offers substantial cost-effectiveness. The evidence in the London experiments in particular suggests that banishing RJC to low-seriousness crimes is a wasted opportunity. If governments wish to fund Restorative Justice at all, this evidence suggests that the best return on investment will be with violent crimes, and with offenders convicted after long prior histories of convictions.

5.2 IMPLICATIONS FOR RESEARCH

The take-up rate by offenders and victims for testing RJC in these experiments was neither low nor high. Had they been higher—upwards of 66% or more—the potential of this method for reducing crime might become more testable. Had they been lower, or under 25%, the potential value of the method might be seen to be reduced. Yet many attempts to introduce RJC run into major difficulties of recruitment and retention of cases. The evidence in this review suggests that perhaps even greater benefits from RJC could be obtained by finding ways to increase the take-up rate. New research could also test ways to increase the delivery rate for RJC when both parties consent. Future research should perhaps focus on the practical issues of delivering high-integrity implementation of RJC. Experiments designed to compare different delivery mechanisms could also include offender and victim outcome measures to add to the evidence on what works in restorative justice.

It is also important for future research to include qualitative measures of the amount of harm that offenders cause before and after they engage in an RJC. The Shapland et al (2008) studies in particular show how this can be accomplished. As new countries, especially in Latin America, attempt to conduct experimental evaluations of RJC, the chance to measure its benefits in this way should not be missed.

One way to interpret these results is to say that the effect of RJC on very serious or frequent offenders was to make them hurt people less. That is just what the empathy-based theory of shared values emerging from effective interaction rituals (Collins, 2004) would predict. Yet it is not possible to observe it by merely counting the numbers of crimes or arrests. The Shapland et al. (2008) innovation in evaluating a crime prevention program is a major success in showing how much difference, and how much more precision, outcome measures based on costs of crime have to offer compared to counts of crimes.

The far greater sensitivity of cost of crime data also means that the smaller sample sizes of experiments testing difficult to implement innovations are not doomed to failure. The low power of counts can be sidestepped by exploiting the great sensitivity of costs. In the process, the cost and difficulty of conducting randomized experiments may potentially be reduced, or their returns on investment may be increased.

DRAFT

6 Acknowledgements

The authors are grateful for support from the Norwegian Knowledge Centre for the Health Sciences, Jerry Lee Center for Criminology, University of Pennsylvania, the Jerry Lee Centre for Experimental Criminology at the University of Cambridge, the Australian National University, the National Institute of Justice, and the Smith Richardson Foundation, not only in preparing this review, but in support for the primary research that underlies much of the content of the reviewed results.

7 Potential conflict of interest

Two of the reviewers conducted most of the studies included in this review. Two of the reviewers are conducting ongoing studies of restorative justice. The reviewers have published other work related to restorative justice.

Two of the ten RCTs were designed, delivered and analyzed by research teams including three of the authors of this review (Strang, Sherman and Woods). Independent authors gathered outcome data, analyzed and published results of the other 8 trials, 7 of which (Shapland et al, 2006, 2008) were operationally directed by two of the authors of this review (Strang and Sherman) and one of which (McGarrell and Hipple, 2007) was operated without contact with any authors of this review. One review author (Sherman) wrote the grant proposals and initial research designs for all ten eligible RCTs. None of the review authors had any conflict of interest in the results of the research, and three of the authors (Sherman, Strang and Woods) conducted the primary research for the only experiment out of ten included in this review that reported a backfiring effect of RJs causing more crime.

8 Contact details for co-reviewers

Lawrence W. Sherman, Ph.D.

Director

Jerry Lee Centre of Experimental Criminology

University of Cambridge

Institute of Criminology

Sidgwick Avenue

CB3 9DA

United Kingdom

Evan Mayo-Wilson, Ph.D.

University College, London, UK

Daniel J Woods, Ph.D

Police Executive Research Forum

Washington, DC, USA

Barak Ariel, Ph.D

University of Cambridge

Institute of Criminology

Cambridge, CB3 9DA, UK

9 References to tests considered for inclusion

Davis, Robert C. (2009). "The Brooklyn Mediation Field Test." Journal of Experimental Criminology 5: 25-39.

McCold P and B Wachtel (1998), The Bethlehem Pennsylvania Police Family Group Conferencing Project, Pipersville, PA: Community Service Foundation. See also <http://www.restorativepractices.org/Pages/bethlehem.html>

McGarrell, E, K Olivares, K Crawford & N Kroovand (2000), Returning Justice to the Community: The Indianapolis Juvenile Restorative Justice Experiment, Indianapolis, Crime Control Policy Center, Hudson Institute. See also <http://www.ncjrs.org/pdffiles1/ojdp/187769.pdf>

McGarrell, Edmund F. (2001) Restorative Justice as an Early Response to Young Offenders. Washington, DC: Office of Juvenile Justice and Delinquency Prevention Bulletin, U.S. Department of Justice.

McGarrell, Edmund F. and Natalie K. Hipple (2007). "Family Group Conferencing and Re-Offending Among First-Time Juvenile Offenders: The Indianapolis Experiment." Justice Quarterly 24: 221-246.

Miers, D., *et al* (2001) An exploratory evaluation of restorative justice schemes. Crime Reduction Research Series, Paper 9. Home Office, Research Development and Statistics Directorate.

Shapland, J., Atkinson, A., Atkinson, H., Chapman, B., Colledge, E., Dignan, J., Howes, M., Johnstone, J., Robinson, G. and Sorsby, A. (2006), Restorative Justice in Practice: The Second Report from the Evaluation of Three Schemes (Sheffield: Centre for Criminological Research, University of Sheffield).

Shapland, J., Atkinson, A. *et al.* (2007). Restorative justice: The views of victims and offenders. The third report from the evaluation of three schemes. Ministry of Justice Research Series 3/07 (London: Ministry of Justice).

Shapland, J., Atkinson, A, Atkinson, H., Dignan, J., Edwards, L., Hibbert, J., Howes, M, Johnstone, J., Robinson, G. and Sorsby, A. (2008), 'Does restorative justice affect reconviction? the fourth report from the evaluation of three schemes' (London: Ministry of Justice).

Shapland, J., G. Robinson & A. Sorsby (2011), Restorative Justice in Practice: Evaluating What Works for Victims and Offenders. London and New York: Routledge

Sherman, Lawrence W. and Heather Strang. (2012). "Restorative Justice as Evidence-Based Sentencing" pp.215-243 in Joan Petersilia and Kevin Reitz, eds., The Oxford Handbook of Sentencing and Corrections. NY: Oxford University Press.

Strang, Heather. (2002). Repair or Revenge: Victims and Restorative Justice. Oxford: Oxford University Press.

Strang, Heather et al 1999-2011 RISE Progress Reports on-line at http://www.aic.gov.au/en/criminal_justice_system/rjustice/rise.aspx

10 Other references

Ahmed, E. (2001). Shame Management Through Reintegration, Melbourne, Cambridge University Press.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.

Angel, C. (2005). 'Victims meet their offenders: testing the impact of restorative justice conferences on victims' post-traumatic stress symptoms' (PhD dissertation, University of Pennsylvania).

Borenstein M, Hedges L, Higgins J, Rothstein H. (2005). *Comprehensive Meta-Analysis Version 2*. Englewood, NJ, Biostat.

Braithwaite J. (1989). *Crime, Shame and Reintegration*. Cambridge, Cambridge University Press.

Braithwaite J. (1998). "Restorative Justice" In Michael Tonry (ed), The Handbook of Crime and Punishment. Oxford: Oxford University Press.

Braithwaite J. (2002). Restorative Justice and Responsive Regulation. Oxford: Oxford University Press.

Bullock, R, Peter van der Laan & Michael Little (1999). Intensive Supervision and Support Programme (ISSP) Kent: A Summary of the Main Outcomes of the Dartington Evaluation. Unpublished report.

Collins, R. (2004). Interaction Ritual Chains. Princeton, Princeton University Press.

Cook, T. and Campbell, D. (1979). Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand-McNally.

Davis, Robert C. (2009). "The Brooklyn Mediation Field Test." Journal of Experimental Criminology 5: 25-39.

DuBourg, R., Hamed, J. and Thorns, J. (2005). 'The economic and social costs of crime against individuals and households'. Home Office Online Report 30/05. (webarchive.nationalarchives.gov.uk/...homeoffice.../rdsolr3005.pdf)

Durkheim, Emile (1912) [2001]. The Elementary Forms of Religious Life: A New Translation by Carol Cosman. Oxford: Oxford University Press.

Eisner, M. (2009). 'No effects in independent prevention trials: can we reject the cynical view?' Journal of Experimental Criminology 5(2) pp163-183.

McCold, P. and Wachtel, B. (1998). Restorative Policing Experiment: The Bethlehem Pennsylvania Police Family Group Conferencing Project. Pipersville, PA: Community Service Foundation.

McGarrell, Edmund F. and Natalie K. Hipple (2007). "Family Group Conferencing and Re-Offending Among First-Time Juvenile Offenders: The Indianapolis Experiment." Justice Quarterly 24: 221-246.

Peachey, D. (1989). The Kitchener Experiment. In M. Wright and B. Galaway (eds), Mediation and Criminal Justice: Victims, Offenders and Community. London: Sage Publications.

Petrosino, A. and Soydan, H. (2005). "The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research." Journal of Experimental Criminology 1(4): 435-450.

Piantadosi, S. Clinical Trials: A Methodologic Perspective, New York, John Wiley & Sons.

Rossner, Meredith (2011). "Emotions and Interaction Ritual: A Micro Analysis of Restorative Justice." British Journal of Criminology 51: 95-119.

Sherman, L. W. (1993). 'Defiance, deterrence and irrelevance: a theory of the criminal sanction', Journal of Research in Crime and Delinquency, 30:445-473.

Sherman, L. W., Strang, H., & Woods, D. J. (2000). Recidivism Patterns in the Canberra Reintegrative Shaming Experiments (RISE). Canberra: Centre for Restorative Justice, Research School of Social Sciences, Australian National University.

Sherman, L. W. & Strang, H., in collaboration with Woods, D. J., Angel, C., Barnes, G. C., Inkpen, N., Newbury-Birch, D., and Bennett, S. B. (2004). Restorative Justice: What we know and how we know it. Philadelphia: University of Pennsylvania, Jerry Lee Center of Criminology, <http://www.sas.upenn.edu/jerrylee/rjWorkingPaper1.pdf>.

Sherman, L. W., H. Strang, C. Angel, D. Woods, G.C. Barnes, S. Bennett & N. Inkpen (2005) 'Effects of face-to-face restorative justice on victims of crime in four randomized, controlled trials', Journal of Experimental Criminology, 1:367-395.

Sherman, L.W., Strang, H., Barnes, G. C. and Woods, D. (2006). 'Race and restorative justice: differential effects for Aboriginals and Whites in the Canberra RISE project.' Los Angeles. Paper presented to the American Society of Criminology.

Sherman, L. & Strang, H. (2007) Restorative Justice: The Evidence, London, Smith Institute.

Sherman, Lawrence and Heather Strang (2011) "Empathy for the Devil: The Nature and Nurture of Revenge" In S. Karstedt, Ian Loader and Heather Strang, eds. Emotions, Crime and Justice. Oxford: Hart Publishing.

Sherman, Lawrence W. and Heather Strang. (2012). "Restorative Justice as Evidence-Based Sentencing" pp.215-243 in Joan Petersilia and Kevin Reitz, eds., The Oxford Handbook of Sentencing and Corrections. NY: Oxford University Press.

Strang, Heather. (2002). Repair or Revenge: Victims and Restorative Justice. Oxford: Oxford University Press.

Strang, H., Barnes, G., Sherman, L., Bennett, S. and Inkpen, N. (2005). 'The effects of restorative justice conferences on Crown Court sentencing of confessed robbers and burglars: a report to the Esmee Fairbairn Foundation.' Philadelphia, PA: Jerry Lee Center of Criminology, University of Pennsylvania.

Strang, H., Sherman, L., Angel, C., Woods, D., Bennett, S., Newbury-Birch, D., and Inkpen, N. (2006) 'Victim evaluations of face-to-face restorative justice conferences: A quasi-experimental analysis.' Journal of Social Issues, 62 (2): 281-306.

Stuart, B. (1996). Circle Sentencing: Turning Swords Into Ploughshares. In B. Galaway and J. Hudson (eds), Restorative Justice: International Perspectives. Monsey, N.Y.: Criminal Justice Press.

Tyler, T. (1990). Why People Obey the Law. New Haven: Yale University Press.

Tyler, Tom and Yuen J. Huo (2002). Trust in the Law: Encouraging Public Cooperation with the Police and Courts. N.Y.: Russell Sage Foundation.

Umbreit, M., Coates, R., Kalanj, B. (1994). Victim Meets Offender: The Impact of Restorative Justice and Mediation. Monsey, N.Y.: Criminal Justice Press; 1994

Weiss, D. & C. Marmar (1997). 'The Impact of Events Scale-Revised' in J. Wilson & T. Keane (eds) Assessing Psychological Trauma and PTSD. New York: Guildford.

Woolf, P. (2009). The Damage Done. London, Bantam Press.

11 Appendix A

11.1 STUDIES EXCLUDED FROM THE REVIEW:

One attempted RCT was excluded from the review, an unpublished report of the Dartington evaluation of the Intensive Supervision and Support Programme in Kent, United Kingdom, where the randomisation was unsuccessful (Bullock *et al.*, 1999).

One RCT was excluded because it did not implement RJs as defined in the review. This experiment, conducted in New York City in the early 1980s, diverted serious offenses from prosecution to Victim-Offender Mediation (Davis, 2009).

Two RCTs of RJs were excluded because they did not include personal victims of crime in the samples. Both of these studies were conducted by two of the authors, Strang and Sherman.

Two studies were excluded because random assignment was performed before consent was obtained. In the Bethlehem Property Experiment (McCold & Wachtel, 1998), which studied juvenile cases of property crimes that generally involved no contact with victims, only 48.6% of offenders were treated as assigned (with the majority refusing to participate in a conference when invited to). In the Bethlehem Violence Experiment (McCold & Wachtel, 1998), which studied juvenile cases of violent crimes, only 31.6% of offenders were treated as assigned (again due to post-random assignment refusals). Both studies included only first-time offenders aged 10-17. In both experiments, cases were randomly assigned either to the control group—prosecution in court—or to diversion to a conference that would leave no criminal conviction record.