

Are Restorative Justice Conferences Effective in Reducing Repeat Offending? Findings from a Campbell Systematic Review

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

Published online: 25 March 2014

© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract

Objectives This paper synthesizes the effects on repeat offending reported in ten eligible randomized trials of face-to-face restorative justice conferences (RJCs) between crime victims, their accused or convicted offenders, and their respective kin and communities.

Methods After an exhaustive search strategy that examined 519 studies that could have been eligible for our rigorous inclusion criteria, we found ten that did. Included studies measured recidivism by 2 years of convictions after random assignment of 1,880 accused or convicted offenders who had consented to meet their consenting victims prior to random assignment, based on “intention-to-treat” analysis.

Results Our meta-analysis found that, on average, RJCs cause a modest but highly cost-effective reduction in the frequency of repeat offending by the consenting offenders randomly assigned to participate in such a conference. A cost-effectiveness estimate for the seven United Kingdom experiments found a ratio of 3.7–8.1 times more benefit in cost of crimes prevented than the cost of delivering RJCs.

Conclusion RJCs are a cost-effective means of reducing frequency of recidivism.

L. W. Sherman · H. Strang (✉) · B. Ariel
Cambridge University, Cambridge, UK
e-mail: hs404@cam.ac.uk

L. W. Sherman
University of Maryland, College Park, MD, USA

H. Strang
Australian National University, Canberra, ACT, Australia

E. Mayo-Wilson
Johns Hopkins University, Baltimore, MD, USA

D. J. Woods
Police Executive Research Forum, Washington, DC, USA

B. Ariel
Hebrew University, Jerusalem, Israel

Keywords Restorative justice · Conferencing · Systematic review · Meta-analysis · Recidivism

Introduction

“Restorative justice” is a contemporary 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: “family group conferences,” the traditional Maori practice which in 1989 became the primary basis for dealing with juvenile crime in New Zealand, “diversionary conferences,” the name used in Australia to describe both juvenile and adult restorative justice as an alternative to prosecution, and “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 RJCs.

Our definition of an RJC is 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 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 officers—trained to organize and convene an RJC that could last from 60 to 180 min or more. The elements of the entire protocol included all of the following:

1. Facilitators conduct a pre-conference screening 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, explaining and obtaining consent to random assignment, which is then used to select those who are actually invited to attend an RJC.
2. Scheduling of a conference at the victims' convenience.
 3. 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.
 4. Introducing all participants in terms of how they are emotionally connected to the crime under discussion.
 5. Opening the discussion by asking offenders to describe the crime they committed.
 6. Inviting victims and all participants to describe the harm the crime has caused and to whom
 7. 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.
 8. Filing the agreement with a court, a police unit, or some other institutional mechanism for monitoring and 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.

The objective of this review is to answer a central question about this practice: what is the effect on repeat offending of a policy of attempting RJC's with consenting victims and offenders?

An equally important question is addressed elsewhere: what are the effects of a policy of attempting RJC's with consenting victims and offenders on various measures of whether victims have been restored to their circumstances prior to the crime? (see Strang 2002; Strang et al. 2006, 2013; Sherman et al. 2005)

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.

Theoretical Basis for Predicting Less Recidivism

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 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 RJCs 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. RJCs 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 RJCs 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.

Review Methodology

This review of the effects of RJCs was limited to studies that had all eight of the following characteristics: (1) reported in the English language; (2) tested a Restorative Justice Conference (RJC) as defined above; (3) 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) offenders in the study had been accused or convicted of committing crimes against one or more identifiable individuals; (5) both 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; (6) study reported data on the frequency of post-random assignment criminal convictions of offenders or re-arrest for 2 years after random assignment; (7) study reported data that enabled the calculation of an intention-to-treat (ITT) effect, rather than treatment as delivered effect; (8) study was conducted after 1994.

These criteria are justified by the following considerations. First, we had no resources for searching in languages other than English. Second, as Braithwaite (1998, 2002) suggests, the restorative justice label embraces a wide range of similar programs that have very different dynamics, but a systematic review of an intervention is most useful when it is focused on a homogeneous protocol. The differences across all things called "restorative justice" could create heterogeneity in the program content that would limit the face validity of our systematic review.

A leading example of what we have omitted is Victim-Offender Mediation (VOM) programs. These have been advocated by two of the most influential figures in the

American RJ movement, Howard Zehr and Mark Umbreit. Zehr, a practitioner and theorist since the emergence of RJ in its modern manifestation in the late 1970s, has worked for a focus on restoration rather than retribution for the benefit of both victims and offenders (see for example Zehr 1990; Zehr and Mika 2003). Mark Umbreit's research into the practice of Victim Offender Mediation (VOM) has also been highly influential in the United States (see for example Umbreit et al. 2004). VOM, however, 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. 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, 2004). 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). Finally, we excluded the Mills et al. (2013) experiment published after the formal search processes primarily on the grounds that it was closer to the VOM model than an RJC, particularly because victims did not have to consent to attend its "Circles of Peace" variant of RJ at all, let alone prior to random assignment.

Third, we require random assignment because it 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.

Fourth, 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 Woolf (2009), 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.

Fifth, 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.

Sixth, 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 ($p = 0.013$), 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 versus 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 et al. 2000; McGarrell 2001; 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 2-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 U.S. Office of Management and Budget.

Seventh, the use of an intention-to-treat (ITT) criterion is, in the authors’ view, essential for this review (Piantadosi 1997: 276–278). 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 RJCs in relation to the total cost of the attempts—including both successes and failures.

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: 2-year post-treatment differences in the frequency of criminal convictions per offender for nine of the studies, and of arrest in Indianapolis.

Eighth, the 1994 threshold reflected the advent of the particular model we identified as most appropriate for a review. Experiments testing RJ programs prior to that date were not based on the Braithwaite-MacDonald orientation of both theory and training. The few

randomized experiments of which we are aware before that date were all based on a VOM model.

We 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 RJC 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.

Results

In all, 15 RCTs and one study that appeared to be an RCT were considered in greatest detail for the review. Six of these were excluded; ten remained. 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 1,880 offenders accepting responsibility for their crimes. The main characteristics of each experiment are described in Table 1.

Assessment of Methodological Quality of the Studies

None of the included studies reported any threats to the integrity of the random assignment process. Randomization was in the hands of the research staff in the Canberra RISE (Reintegrative Shaming) Experiments (nos. 1, 2 in Table 1) 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 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 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.

Table 1 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 %	248
3. Indianapolis	1997 ^a	McGarrell and Hipple	Violence and property under 14	Diversion from prosecution	Other diversion programs; VOM	80 %	782
4. London	2001–2005	Shapland et al.	Robbery over 18	Post-plea, presentence	No RJC presentence	85 %	106
5. London	2001–2005	Shapland et al.	Burglary over 18	Post-plea, presentence	No RJC presentence	85 %	186
6. Thames Valley	2001–2005	Shapland et al.	Assault over 18	In prison, prerelease	No RJC in prison	73 %	103
7. Thames Valley	2001–2005	Shapland et al.	Assault over 18	On probation	No RJC on probation	87 %	63
8. Northumbria	2001–2005	Shapland et al.	Violence and property under 18	Diversion to a police caution with RJC	Diversion to caution with no RJC	92 %	165
9. Northumbria	2001–2005	Shapland et al.	Property over 18	Post-plea, presentence	No RJC presentence	77 %	105 ^b
10. Northumbria	2001–2005	Shapland et al.	Violence over 18	Post-plea, presentence	No RJC presentence	(77 %)	(105)
Total							1,880

^a McGarrell and Hipple do not report the date on which they stopped random assignment

^b 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

As Table 1 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 was reported by Shapland et al. (2006: 25) to vary between a low of 73 % for the Thames Valley Prison experiment and a high of 92 % for the Northumbria youth 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: 276–278). 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 1 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 1 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.

In most of the ten experiments, imprisonment was rarely used in either the RJC or control group cases (though in the case of experiment #6, the offenders were 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 et al. (2008) was to eliminate randomly assigned cases from the analysis if the offenders had served the entire 2 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 2-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.

Meta-analysis of Repeat Offending Effects

The primary criterion of the effect of RJC on crime is the frequency of repeat offending over the 2 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 Fig. 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

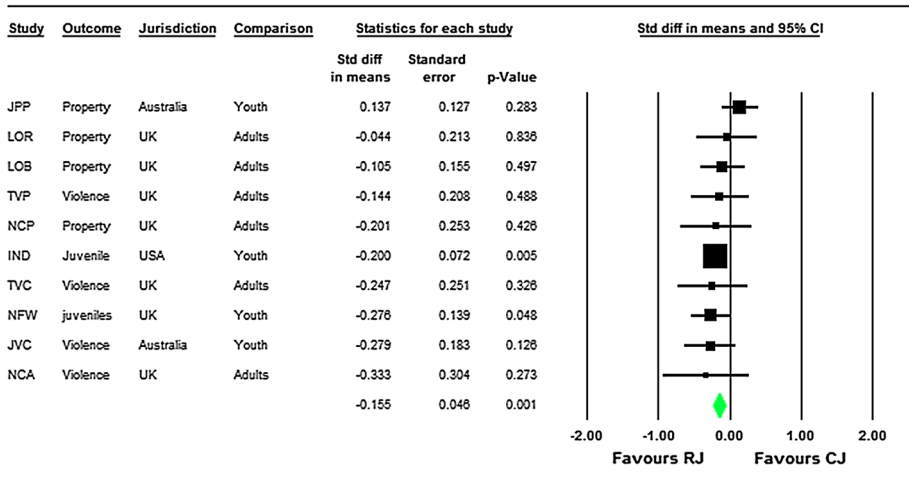


Fig. 1 Combined effects sizes for study outcomes. Meta-analysis random effects model, $Q = 7.754$, $df = 9$, $p < 0.559$

Figure 1 shows that the average effect of RJC is to reduce crime. More precisely, across 1,880 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 ($d = .155$, $p = 0.001$).

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 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 (but not significantly more)—the *pattern* of findings can be described as statistically “significant” and favoring the benefits of RJC. That means, in this case, that there is only a one in a thousand chance that the pattern in Fig. 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 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 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 Fig. 1.

Moderator Analyses

The overall meta-analysis of the ten experiments can be unpacked to learn whether RJC 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” we mean anything other than the independent and dependent variables

(RJC and repeat convictions) that could affect or “moderate” the relationship between those two variables. In this review we limited largely to intentional differences in design of each experiment, all of which amount to testing RJC with different kinds of people, of crime types, or of stages of the criminal justice process. The only third factor “by design” that we ignored was the nation in which the experiment was conducted, since we had no clear theoretical basis for separating it from the more fundamental issue of using RJC as diversion from criminal justice (as in all Australian and US tests) or supplementation to CJ (as in all but one English test).

Differences in mean effects by moderator variables could suggest, for example, that if RJC were used with only the kinds of cases associated with that 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 Figs. 2, 3. The three moderators for which we had adequate power to make comparisons were violent versus property crime, juvenile versus adult offenders, and use of RJC as diversion from or a supplement to criminal justice.

Violent Versus Property Crime

Half of the experiments in the sample tested RJC with violent crimes. Figure 2 shows what the average effect of RJC 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 ($Q = 1.021, p = 0.9$) 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

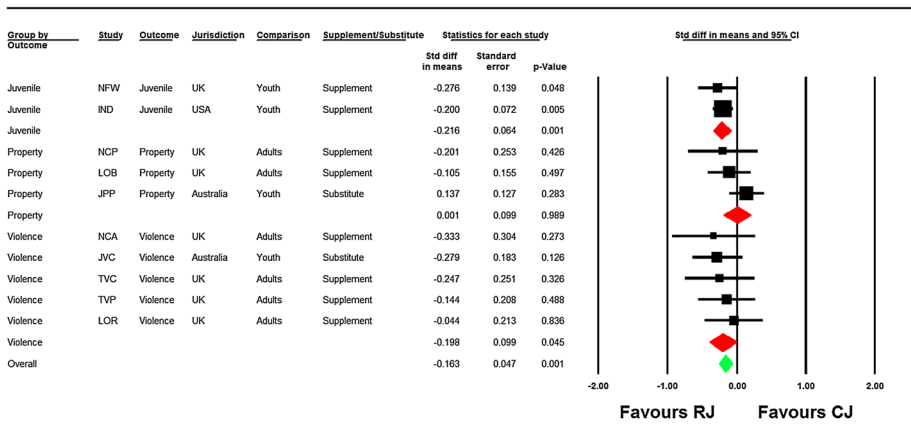


Fig. 2 Crime type as moderator of study outcomes. Juveniles $Q = 0.233, df = 1, p < 0.630$; property $Q = 2.244, df = 2, p < 0.326$; violence $Q = 1.021; df = 4, p < 0.907$; between group $Q = 3.574 df = 2, p < 0.167$

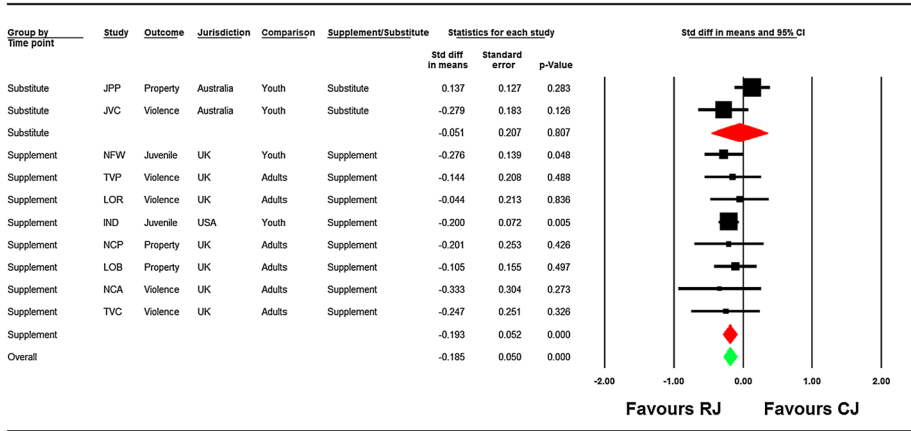


Fig. 3 Effects of RJC as supplement or substitute to conventional justice on frequency of repeat offending, 2-year follow-up period. RJC as substitute $Q = 3.491, df = 1, p < 0.062$; supplement $Q = 1.483; df = 7, p < 0.983$; between group $Q = 0.447; df = 1, p < 0.504$

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.

Juvenile Versus Adult Crime

Many public officials say that RJC is more appropriate for juvenile offenders than for adults. Yet the findings from this Review suggest otherwise, at least for offenses with personal victims. In Fig. 2 we see that the average effect of RJC in six experiments with all adults is .150 standard deviations fewer future convictions than without RJC. 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. Nonetheless, it is in the opposite direction from the conventional wisdom.

Diversion Versus Supplementation

One of the major policy debates in restorative justice is whether it should merely supplement conventional justice (CJ), or replace it altogether. In Fig. 3 we see that the average effect of RJC is larger in the eight 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 Fig. 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.

Sensitivity Analyses

A sensitivity analysis is a kind of moderator that is not necessarily by design, but still allows comparisons across mean effects of several experiments in different categories to determine whether effects are sensitive to those categories. 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 (1) effects of the authors as evaluators, (2) the effects of using arrests (in Indianapolis) in a meta-analysis that uses convictions in all nine other experiments, and (3) 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 2 years after random assignment (or treatment).

Authors as Evaluators

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 (see Shapland et al. 2006, 2007, 2008, 2011).

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 Fig. 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.

Arrests Versus 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

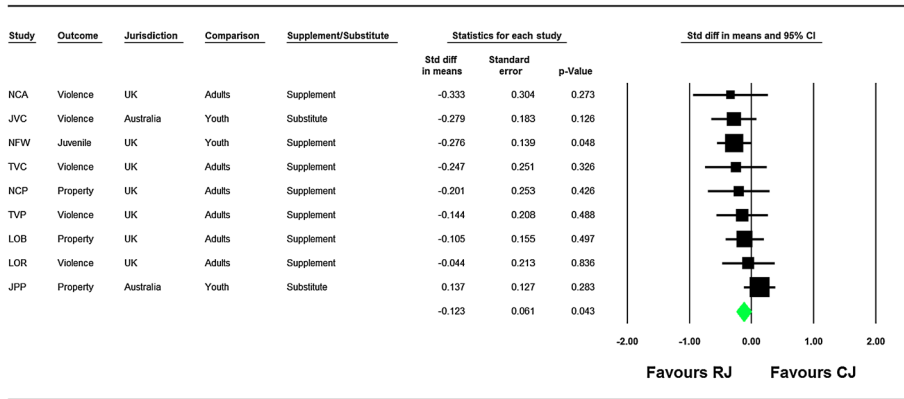


Fig. 4 Effects of RJC on the frequency of criminal convictions, 2-year post-treatment follow-up period

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.

Time at Risk Out of Prison

Another difference across experiments is the inclusion of time-at-risk periods when offenders were in prison in some tests but not in others. As noted above, 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 2 years after date of random assignment, from one day to 2 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 2-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 2-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 Fig. 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 2-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

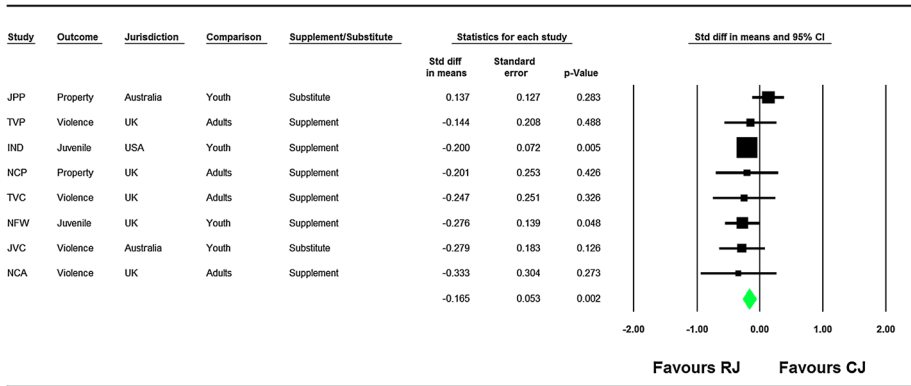


Fig. 5 Effects of RJC on the frequency of repeat offending (without deletions for time at risk), 2-year post-treatment follow-up period



Fig. 6 Effects of RJC on the frequency of repeat offending (with deletions for no time at risk), 2-year post-treatment follow-up period

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 et al. (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.

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 with equal weight and importance. 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 (Sherman 2011). 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 (Sherman 2013: 422–425).

In seven of the experiments included in this review, the primary evaluators (Shapland et al. 2008) took the highly original and very important step of giving widely varying weight to each crime for which offenders were convicted in the 2-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 10–1. 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 et al. (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 et al. (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.

Table 2 presents the most specific data available on both costs of crimes prevented and recidivism counts. While the recidivism is available by experiment, the costs of crime (and of restorative justice) are only available by site (Shapland et al. 2008:64). The reason for that was the difficulty of distinguishing costs for each experiment, since each site simultaneously spread salary and other costs across multiple experiments. What Table 2 shows is the total cost of RJC and total costs of crimes prevented relative to the cost of crimes in the control groups’ 2-year followup period for the entire site, with the relative frequency contrasts of convictions presented by experiment in the far right-hand column.

The most striking finding in Table 2 is how the effect size of the impact assessment can be changed substantially by using costs of crimes rather than raw counts unadjusted for harm levels. While the experiments with robbery and burglary offenders in London yielded small and non-significant effect sizes of RJC on the frequency of reconstructions, the cost-effectiveness ratio in London was £3.7 in costs of crime prevented for every £1 spent on delivering RJC. As compared to the 61 % lower reconstruction frequency of violent offenders given RJC in the Northumbria Magistrates’ court (Sherman and Strang 2012: 231), the London robbery experiment had only 8 % fewer reconstructions and the London burglary experiment had only 16 % fewer reconstructions for RJC cases than controls. But when the cost-effectiveness ratio of RJC in London was compared to the cost effectiveness ratio in Northumbria, they were much closer in size: 5.1–1 in Northumbria across all 3 experiments compared to the 3.7–1 in the two London experiments. All of the cost

Table 2 Cost effectiveness versus recidivism reduction in seven tests

Experimental name	Running costs of RJC	Cost of crime estimates for crime prevented by RJC	Cost-effectiveness ratio	Relative percent reduction in frequency of reconvictions, RJC versus controls, 2-year after-only
London Robbery and Burglary (2)	£598,848	£2,214,811	3.7–1	8 % robbery 16 % burglary
Northumbria Juvenile, Adult Violence and Adult Property Crime (3)	£275,411	£1,414,593	5.1–1	48 % juveniles 61 % adult assaults 24 % Adult Property Crime
Thames Valley Prison and Probation for Violence (2)	£222,463	£1,808,952	8.1–1	55 % probation 33 % prison

Cost of crime and RJC's data all taken from Shapland et al. (2008) at page 64, using "Method 2" explained on page 62, which corresponds to the 2 years, after-only comparison of experimental and control cases which is the same basis as this review uses for comparing frequency of reconvictions, and which is used elsewhere in Shapland to estimate recidivism effects of English experiments. The cost of crime estimates in Shapland et al. (2008) are derived from Home Office standard values based on Dubourg et al. (2005). The present table reports different findings from those presented in Figure 9.3 in Sherman and Strang (2012: 232) because the 2012 figure uses Shapland et al's (2008: 64) "Method 3," a before–after comparison difference of differences between experimental and control cases. That before–after design does not match the after-only difference design for all the analysis in the present article, to which the present table conforms

calculations in the UK sites were statistically significant, even where they were not significant for comparing counts of crimes. The biggest cost-effectiveness ratio was in the post-conviction experiments for violent crimes in *Thames Valley*, where the combined probation and prison experiments yielded £8 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 can find the necessary 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. But they must also concede that the Shapland et al. (2008) analysis was unable to estimate the direct benefits of RJC effects on the crime victims who experienced RJC; also had benefits that likely saved the National Health Service money, as documented in the work of Angel et al. (2014); the analysis was limited only to costs of recidivism, which understate total benefits of RJC for the same cost invested.

Conclusions and Implications

Restorative justice conferences 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 RJC, 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 RJC without consent are arguably unethical. The Review’s conclusions are appropriately limited to the kinds of cases in which RJC would be ethical and appropriate. Among the kinds of cases in which both offenders and victims are willing to meet, RJC seem likely to reduce frequency and (with less data) costs of future crime.

Implications for Criminological Theory

These findings suggest that RJC may be a means by which authorities can foster “turning points” (Sampson and Laub 1991; Laub and Sampson 2001) in criminal careers. If

desistance is defined as a process rather than as a binary change, the reduced frequency and cost of the crimes committed after offenders participate in an RJC is consistent with the conference triggering a process of desistance.

The causal pathway from the conference to desistance over the next 2 years is consistent with Collins (2004) theory of “interaction ritual chains,” as an analysis of systematic observation data in the Australian trials suggests (Rossner 2013). That study’s correlational analyses within the RJC groups showed that the more intensive and meaningful the quality of what Durkheim called “collective effervescence” in a highly emotional discussion of the harm a crime had caused to someone in the room, the bigger the effect seems to be on reducing recidivism. While it is difficult to imagine the intensity of interaction ritual could be randomly assigned, it is clear that there is more intensive ritual in RJC than in conventional court processes (Rossner 2013). That difference makes collective effervescence one causal pathway consistent with these findings, if not the only possible explanation. If it is a precondition of success in fostering desistance, this theoretical interpretation gives the evidence even clearer implications for practice.

Further evidence in support of this theoretical mechanism, as one of our reviewers for this article pointed out, comes from the comparison of effect sizes for tests on violent crime cases versus tests on property crimes. The larger effect sizes for violent crimes is a finding consistent with the greater emotion usually associated with discussions of violent crimes, compared to discussing vandalism, car break-ins or even burglary. While further research could systematically compare indicators of interaction ritual strength in violent versus property crime conferences, the extent of harm and length of prison sentences are consistent with more intensive interactions, on average, with violent crimes, and consistent with the greater effectiveness of RJC with those cases.

Implications for Practice

The effects of RJC 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 RJC 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 RJC as a supplement. Cost-saving goals have apparently strengthened the appeal of RJ in theory, but without the kind of evidence reviewed here.

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 UK evidence on using RJC as a supplement also offers substantial cost-effectiveness as well as reductions in recidivism. We cannot yet say the same about the cost-effectiveness of RJ as a diversion, since a similar cost of crime analysis has not yet been done for the Australian experiments (although a 15-year followup analysis of the cost-effectiveness of RJC is underway).

Until we have more comprehensive data on the diversion-supplementation issue, we can draw our best global lessons from the violence versus property moderator analysis. If governments wish to fund Restorative Justice at all, this evidence suggests that the best return on investment will be with violent crimes. This evidence makes sense theoretically, because of the more intense interaction ritual that is likely to emerge in an RJC on violent crimes compared to physically injurious crime. We have seen, for example, an RJC for a London robbery of a cab driver that injured him so badly he was in the hospital for

2 weeks. The offender who admitted the crime was in tears for most of a 2-h conference; many other RJC's with violent crime also provoked tears and other signs of strong emotions by various participants. We can recall no RJC for a property crime which had no contact between victim and offender in which such strong emotions were evoked. Interviews of London officers who led both kinds of conferences confirm that the conferences about more serious physical injuries evoked more intense emotions (Rossner 2014).

Perhaps the most important implication for practice is that these results must be seen as the effects of a very specific, highly homogenous model of restorative justice. There is no basis at all for generalizing the conclusions from this review to other models of RJ practice. In the UK, for example, thousands of minor crimes annually are now dealt with on the street as on-the-spot, police-led street encounters called "community resolutions." These methods are based on restorative principles, but are far too brief and public to meet the standards of interaction ritual theory. Readers should be well-advised that nothing in the present review provides any evidence in support of the claim that these brief events reduce crime or help victims. Moreover, the review cannot be extended to the many other models of RJ practice that remain untested by randomized controlled trials.

This warning does not mean that the review shows that quick-fixes or other RJ approaches cannot work. It simply means that the time-consuming preparations for a 2–3 h 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. People skilled in practice may often wish to generalize beyond the data to justify what they wish to do. But it is not appropriate to call any other form of RJ "evidence-based" in the way that medical practices are defined as evidence-based: repeated randomized trials with meta-analysis. The good news, however, is that this review can be said to move this kind of RJ practice into a legitimate classification as "evidence-based."

Implications for Research

The rate at which offenders and victims consent to participate in testing RJC's 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 rates (as distinct from individual recidivism) might become more testable. Had they been lower, or under 25 %, the potential value of the method for reducing crime rates might be seen to be reduced. Yet many attempts to introduce RJC's run into major difficulties of recruitment and retention of cases. The evidence in this review suggests that perhaps even greater benefits from RJC's could be obtained by finding ways to increase the "takeup" (consent) rate among victims and offenders.

New research could also test ways to increase the delivery rate for RJC's when both parties consent. Future research should perhaps focus on the practical issues of delivering high-integrity implementation of both RJC's and routine practice. 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. For example, an RCT could compare the cost-effectiveness of contacting victims by telephone or meeting with them in their homes, face-to-face, with the outcome being whether the victim consents to participate. Even predictive research modeling which case characteristics increase the likelihood of offenders and victims consenting to RJ—comparable to "solvability factors" in investigations—or what might be called "consentability." Those models might even

change over time if RJC's become more widespread and more victims and offenders would have heard about them prior to a request for their consent to participate in one.

It is also important for future research to include quantitative and 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, using the Home Office cost of crime estimates (Dubourg et al. 2005). As new countries attempt to conduct experimental evaluations of RJC's, the chance to measure its benefits in this way should not be missed. At minimum, the Crime Severity Index used by the Canadian government can provide a weighting of each crime type based on the prescribed length of prison sentence for each offense of that type.

The value of cost of crime data is apparent too from the success of the Shapland et al. (2008) innovation in showing how much difference, and how much more precision, outcome measures based on costs have to offer, compared with counts of crime. The far greater sensitivity of cost of crime data also means that the smaller sample sizes of experiments testing such difficult-to-implement innovations are not doomed to failure. The low power of counts can be sidestepped by exploiting the great sensitivity of costs, or of crime severity measured by sentencing guidelines. In the process, the cost and difficulty of conducting randomized experiments may potentially be reduced, and their returns on investment may be increased.

Acknowledgments This article is based on a Campbell Collaboration Systematic Review (Strang et al. 2013). The authors are grateful for support from the Norwegian Knowledge Centre for the Health Sciences, the Jerry Lee Center for Criminology at the 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. This article was prepared at the Regulatory Institutions Network at the Australian National University, School of Pacific and Asian Studies, with benefit of a stipend from the Norwegian Knowledge Centre for the Health Sciences. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice or the Norwegian Knowledge Centre for the Health Sciences.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Ahmed E (2001) Shame management through reintegration. Cambridge University Press, Melbourne
- Angel C, Sherman L, Strang H, Ariel B, Bennett S, Inkpen N, Keane A, Richmond T (2014) Short-term effects of restorative justice conferences on post-traumatic stress symptoms among robbery and burglary victims: a randomized controlled trial. *J Exp Criminol* (forthcoming)
- Borenstein M, Hedges L, Higgins J, Rothstein H (2005) Comprehensive meta-analysis version 2. Biostat, Engelwood
- Braithwaite J (1989) Crime, shame and reintegration. Cambridge University Press, Cambridge
- Braithwaite J (1998) Restorative justice. In: Tonry M (ed) *The handbook of crime and punishment*. Oxford University Press, Oxford
- Braithwaite J (2002) Restorative justice and responsive regulation. Oxford University Press, Oxford
- Collins R (2004) Interaction ritual chains. Princeton University Press, Princeton
- Cook T, Campbell D (1979) Quasi-experimentation: design and analysis issues for field settings. Rand-McNally, Chicago
- DuBourg R, Hamed J, 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 E (1912) [2001] *The elementary forms of religious life: a new translation* by Carol Cosman. Oxford University Press, Oxford
- Eisner M (2009) No effects in independent prevention trials: can we reject the cynical view? *J Exp Criminol* 5(2):163–183
- Laub JH, Sampson RJ (2001) Understanding desistance from crime. *Crime Justice Rev Res* 28:1–69
- McCold P, Wachtel B (1998) The Bethlehem Pennsylvania police family group conferencing project. Community Service Foundation, Pipersville, PA. See also <http://www.restorativepractices.org/Pages/bethlehem.html>
- McGarrell EF (2001) Restorative justice as an early response to young offenders. Office of Juvenile Justice and Delinquency Prevention Bulletin, U.S. Department of Justice, Washington, DC
- McGarrell EF, Hipple NK (2007) Family group conferencing and re-offending among first-time juvenile offenders: the Indianapolis experiment. *Justice Q* 24:221–246
- McGarrell E, Olivares K, Crawford K, Kroovand N (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>
- 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
- Mills L, Barocas B, Ariel B (2013) The next generation of court-mandated violence treatment: a comparison study of batterers' treatment and restorative justice programs. *J Exp Criminol* 9:65–90
- Peachey D (1989) The Kitchener experiment. In: Wright M, Galaway B (eds) *Mediation and criminal justice: victims, offenders and community*. Sage Publications, London
- Petrosino A, Soydan H (2005) The impact of program developers as evaluators on criminal recidivism: results from meta-analyses of experimental and quasi-experimental research. *J Exp Criminol* 1(4):435–450
- Piantadosi S (1997) *Clinical trials: a methodologic perspective*. Wiley, New York
- Rossner M (2011) Emotions and interaction ritual: a micro analysis of restorative justice. *Brit J Criminol* 51:95–119
- Rossner M (2013) Just emotions: rituals of restorative justice. *Clarendon studies in criminology*. Clarendon Press, Oxford
- Rossner M (2014) Personal communication with second author, 15 February
- Sampson RJ, Laub JH (1991) *Crime in the making*. Harvard University Press, Cambridge
- Shapland J, Atkinson A, Atkinson H, Chapman B, Colledge E, Dignan J, Howes M, Johnstone J, Robinson G, Sorsby A (2006) *Restorative justice in practice: the second report from the evaluation of three schemes*. Centre for Criminological Research, University of Sheffield, 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. Ministry of Justice, London
- Shapland J, Atkinson A, Atkinson H, Dignan J, Edwards L, Hibbert J, Howes M, Johnstone J, Robinson G, Sorsby A (2008) *Does restorative justice affect reconviction? The fourth report from the evaluation of three schemes*. Ministry of Justice, London
- Shapland J, Robinson G, Sorsby A (2011) *Restorative justice in practice: evaluating what works for victims and offenders*. Routledge, London
- Sherman LW (1993) Defiance, deterrence and irrelevance: a theory of the criminal sanction. *J Res Crime Delinq* 30:445–473
- Sherman LW (2011) Al Capone, the sword of Damocles, and the police-corrections budget ratio. *Criminol Public Policy* 10:195–206
- Sherman LW (2013) The rise of evidence-based policing: targeting, testing and tracking. In: Tonry M (ed) *Crime and justice in America 1975–2025*. Crime and justice: a review of research, vol 42. University of Chicago Press, Chicago, pp 377–451
- Sherman L, Strang H (2007) *Restorative justice: the evidence*. Smith Institute, London
- Sherman L, Strang H (2011) Empathy for the devil: the nature and nurture of revenge. In: Karstedt S, Loader I, Strang H (eds) *Emotions, crime and justice*. Hart Publishing, Oxford
- Sherman LW, Strang H (2012) Restorative justice as evidence-based sentencing. In: Petersilia J, Reitz K (eds) *The Oxford handbook of sentencing and corrections*. Oxford University Press, NY, pp 215–243
- Sherman LW, Strang H, Woods DJ (2000) *Recidivism patterns in the Canberra reintegrative shaming experiments (RISE)*. Centre for Restorative Justice, Research School of Social Sciences, Australian National University, Canberra
- Sherman LW, Strang H, Angel C, Woods D, Barnes GC, Bennett S, Inkpen N (2005) Effects of face-to-face restorative justice on victims of crime in four randomized, controlled trials. *J Exp Criminol* 1:367–395

- Sherman LW, Strang H, Barnes GC, Woods D (2006) Race and restorative justice: differential effects for Aboriginals and 5 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
- Strang H et al (1999–2011) RISE progress reports on-line at http://www.aic.gov.au/en/criminal_justice_system/rjustice/rise.aspx
- Strang H (2002) Repair or revenge: victims and restorative justice. Oxford University Press, Oxford
- Strang H, Barnes G, Sherman L, Bennett S, 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. Jerry Lee Center of Criminology, University of Pennsylvania, Philadelphia
- Strang H, Sherman L, Angel C, Woods D, Bennett S, Newbury-Birch D, Inkpen N (2006) Victim evaluations of face-to-face restorative justice conferences: a quasi-experimental analysis. *J Soc Issues* 62(2):281–306
- Strang H, Sherman L, Mayo-Wilson E, Woods D, Ariel B (2013) Restorative justice conferencing (RJC). Using face-to-face meetings of offenders and victims: effects on offender recidivism and victim satisfaction. A systematic review. *Campbell Syst Rev* 12
- Stuart B (1996) Circle sentencing: turning swords into ploughshares. In: Galaway B, Hudson J (eds) *Restorative justice: international perspectives*. Criminal Justice Press, Monsey
- Tyler T (1990) *Why people obey the law*. Yale University Press, New Haven
- Tyler T, Huo YJ (2002) *Trust in the law: encouraging public cooperation with the police and courts*. Russell Sage Foundation, New York
- Umbreit M, Coates R, Kalanj B (1994) *Victim meets offender: the impact of restorative justice and mediation*. Criminal Justice Press, Monsey
- Umbreit M, Coates R, Vos B (2004) Victim offender mediation: three decades of practice and research. *Confl Resolut Q* 22(1–2):279–303
- Woolf P (2009) *The damage done*. Bantam Press, London
- Zehr H (1990) *Changing lenses: a new focus for crime and justice*. Herald Press, Scottsdale
- Zehr H, Mika H (2003) Fundamental concepts of restorative justice. In: McLaughlin E (ed) *Restorative justice: critical issues*. Sage Publications, London