



The effectiveness of restorative justice programs: A meta-analysis of recidivism and other relevant outcomes

Criminology & Criminal Justice

1–27

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/17488958231215228

journals.sagepub.com/home/crj**Lindsay Fulham**

York University, Canada

Julie Blais 

Dalhousie University, Canada

Tanya Ruge

Independent Researcher

Elizabeth A Schultheis

Carleton University, Canada

Abstract

Restorative justice is an alternative approach to traditional legal system programming that focuses on repairing harm and enhancing client accountability. Despite the proliferation of restorative justice programs, research suggests that their effectiveness depends on various factors such as program type and methodological quality of the studies. The goal of this study was to synthesize the research on the effects of restorative justice in reducing recidivism as well as improving other outcomes for male and female adult clients. Information from 27 studies examining 34 unique samples was included in the meta-analysis. Results indicated that restorative justice programs were associated with significant and small reductions in general recidivism but not violent recidivism. In addition, restorative justice programs resulted in greater victim and client satisfaction, victims' views of procedural justice, and client accountability compared to traditional legal system approaches. There were significant sample, study, and program moderators that influenced the

Corresponding author:

Julie Blais, Department of Psychology and Neuroscience, Life Sciences Centre, Dalhousie University, 1355 Oxford Street, P.O. Box 15000, Halifax, NS B3H 4R2, Canada.

Email: julie.blais@dal.ca

effects of restorative justice in reducing recidivism outcomes. Taken together, the results over the past 40 years of research provide minimal support for the effectiveness of restorative justice programs in reducing recidivism overall and highlight the importance of considering moderating factors when evaluating and improving the effectiveness of restorative justice programs. Notably, this meta-analysis clearly demonstrates restorative justice's effectiveness in improving other relevant outcomes for clients and victims.

Keywords

General recidivism, meta-analysis, restorative justice, satisfaction, violent recidivism

Restorative justice (RJ) is an alternative approach to resolving conflict that is increasingly being integrated into criminal legal systems around the world (Federal-Provincial-Territorial Working Group on Restorative Justice, 2020; United Nations Office on Drugs and Crime, 2020). In contrast to traditional legal system procedures that emphasize segregation and punishment, the goal of RJ is to repair broken relationships and reintegrate clients into their communities (Daly, 2001). However, despite the movement toward restorative approaches, empirical research supporting their effectiveness in reducing recidivism and improving other outcomes such as victim and client satisfaction is not well established (Gang et al., 2021). According to previous reviews, the mixed findings in the literature may be partly attributed to differences in methodological quality of the studies (Sherman et al., 2015; Wong et al., 2016). In particular, studies that include more rigorous research methods (e.g. randomized controlled trials (RCTs), large sample sizes) are less likely to demonstrate significant reductions in recidivism compared to studies with less rigorous designs. In addition, the wide variation in the principles, techniques, and goals of RJ programs present unique challenges to summarizing RJ's overall effectiveness. The purpose of this study was to provide an updated meta-analysis on the effectiveness of RJ programs in reducing recidivism in adult clients. Given that the goals of RJ extend beyond reducing recidivism, this analysis also evaluated the effectiveness of RJ programs in improving other outcomes such as victim and client satisfaction, victim perceptions of procedural justice, and client accountability.

Defining restorative justice

In general, RJ refers to a variety of practices that focus on remedying the harm (where possible) experienced by those directly affected by the criminal act (Bazemore and Schiff, 2001). One of the most widely cited definitions by Marshall (1996) states that "restorative justice is a process whereby all the parties with a stake in a particular offence come together to resolve collectively how to deal with the aftermath of the offence and its implications for the future" (p. 37; cf. Braithwaite, 1999: 5). The most common types of RJ programs include circles, family-group conferences, and victim–client mediation (Department of Justice Canada, 2018; Piggott and Wood, 2018). While a discussion of each type of program is beyond the scope of this article, interested readers can see Maxwell and Morris (2003), Bazemore and Umbreit (1999), and McCold and Wachtel (2012). Despite the variability across programs, RJ is often a voluntary procedure that

involves bringing together the client, victim, and community in discussing the causes and repercussions of the criminal act. In doing so, RJ allows the opportunity for victims' voices and needs to be heard and empowers them to play an active role in the reparative process. In addition, RJ aims to establish a greater sense of accountability on the client while reducing stigma and reliance on punitive measures (Umbreit, 2001).

RJ is both a theory of criminal justice and an approach to rehabilitation purported to reduce future criminality (Shapland et al., 2008). RJ encompasses theories such as conflicts as property (community reclaiming power in handling criminal matters; Dzur and Olson, 2004), procedural justice (treating clients with respect leading to reciprocal prosocial behavior; Tyler, 1990), and reintegrative shaming (social disapproval as promoting desistance; Braithwaite, 1999). It is unclear where RJ stands in relation to one of the most empirically supported theories of correctional rehabilitation, namely Risk-Need-Responsivity (RNR; Andrews et al., 1990; Andrews and Bonta, 2010). RNR theory encompasses three principles that guide rehabilitation practices: (1) the risk level of the client should be matched with the intensity and duration of the treatment services that they receive (Risk), (2) treatments are most effective when they target criminogenic needs (i.e. factors that are found to influence recidivism risk such as antisocial peers, substance use, and impulsivity; Need), and (3) treatment should utilize empirically supported intervention strategies (i.e. cognitive-behavioral therapy; CBT) while taking into account the abilities and learning styles of individuals (Responsivity; Andrews and Bonta, 2010).

While there is a large body of evidence supporting RNR (e.g. Andrews et al., 2011; Bonta et al., 2010; Looman and Abracen, 2013), the research investigating the application of RNR principles to RJ programs is limited. A major point of disagreement appears to be in RJ's emphasis on the processes that eventually lead to desistance (e.g. healing) rather than targeting specific risk factors. Proponents of RNR have suggested that RJ programs should incorporate RNR principles in order to enhance treatment outcomes (Bonta et al., 2006). Others view RJ and traditional rehabilitation programs as separate, but complimentary approaches (Ward et al., 2014). From this perspective, RJ is both a program that focuses on repairing harm and a set of ethical principles to guide complimentary rehabilitation approaches (Gavrielides and Worth, 2014). Others have suggested that RJ programs and traditional rehabilitation approaches cannot co-exist partly due to the fact that the victim is not considered in traditional rehabilitation (Zernova, 2009). Clearly, a consensus on this issue has yet to be established.

Summary of the RJ literature

One of the major goals of RJ is to repair harms and address participant needs. It is generally believed that RJ leads to greater satisfaction among victims and clients compared to routine criminal justice services (Umbreit, 1998), and there does appear to be some empirical evidence supporting this claim (Latimer et al., 2005; Umbreit et al., 2004). RJ programs may also increase client satisfaction (McGarrell and Hipple, 2007; Umbreit and Coates, 1992). McGarrell and Hipple (2007) reported that juvenile clients were more likely to say that they were treated with respect (97% vs 58%), were more engaged in the program (76% vs 20%), and were more comfortable with expressing themselves (66% vs 24%), compared to

a control group of offending youth. Despite the positive effects of RJ on victim and client satisfaction, RJ may not be appropriate for all victims depending on their levels of distress and the nature of the crime.

Several meta-analyses have been conducted examining the effectiveness of RJ in reducing different recidivism outcomes. Overall, these studies support the effectiveness of RJ programs in reducing recidivism in both juvenile and adult clients (e.g. Bonta et al., 2006; Bradshaw et al., 2006; Latimer et al., 2005); however, several moderators are also noted. For example, studies with more rigorous methodologies tend to show significantly smaller positive effects of RJ compared to traditional correctional programming (Wong et al., 2016). Sample variables, such as the racial composition of the sample, age of the sample (i.e. juvenile vs adult), and type of crime (e.g. violent vs property offenses) have also been shown to moderate the effects of RJ on recidivism (Bain, 2012; Sherman et al., 2015; Wong et al., 2016).

Meta-analyses of RJ programs have also tended to be too narrow in their definition of RJ (examining only VOM, for example; Bradshaw et al., 2006) or too broad, collapsing across different RJ programs without considering program differences as potential moderators (e.g. Latimer et al., 2005). Meta-analyses have also ignored indicators of program fidelity (e.g. extent to which program adhered to RJ principles) and integrity (e.g. organized and structured program). Furthermore, few meta-analyses have taken into consideration important client characteristics that may affect RJ outcomes such as client risk level. Finally, although recidivism is not considered the primary goal of RJ, previous meta-analyses have mostly focused on recidivism as the primary outcome without consideration for other outcomes such as client and victim satisfaction.

Current meta-analysis

This study sought to address the inconsistent results in the literature by conducting a meta-analysis on the effectiveness of RJ in reducing recidivism and improving other outcomes for adult clients, while updating and addressing limitations of previous meta-analyses (e.g. Bonta et al., 2006; Latimer et al., 2005). While this study examined a number of different RJ programs, specific program features were coded to examine their potential moderating effects on the outcomes. RJ programs were also scored on different elements of adherence or “fidelity” to RJ principles (e.g. voluntariness, community representation, healing, reintegration) and on different elements of integrity (e.g. proper training, supervision). More complete information on client characteristics including risk level was also extracted from the included studies. In terms of methodological rigor, a standardized assessment tool was used to systematically assess studies on their methodological quality. Finally, while previous meta-analyses have focused on recidivism as the primary outcome this study took into consideration all possible outcome variables that had a sufficient number of studies to meta-analyze.

Hypotheses

Based on the literature summarized, this study tested four main hypotheses. First, it was hypothesized that RJ would demonstrate greater reductions in recidivism compared to

traditional legal system responses. Second, it was anticipated that RJ would lead to improvements in other outcomes such as victim and client satisfaction, procedural justice, and client accountability. Third, as previous meta-analyses have suggested, the effects of RJ on these outcomes were expected to be contingent on important moderators. Specifically, it was hypothesized that programs that adhered to RJ principles (i.e. fidelity) and had higher program integrity would show stronger reductions in recidivism and improve other outcomes compared to traditional methods. The influence of other program factors, such as type of RJ program (e.g. conferences), was examined for exploratory purposes. Finally, considering research demonstrating study quality as an important moderator, it was hypothesized that more rigorous research designs would yield smaller effect sizes compared to less rigorous research designs using a more comprehensive assessment of methodological quality.

Method

Systematic review

The preferred reporting items for systematic reviews and meta-analyses (PRISMA, Page et al., 2021) statement and the specific checklist for meta-analyses were used to guide the design of this study (e.g. search strategy, selection of studies, data extraction). The literature search was conducted as part of a larger meta-analysis exploring the effects of RJ on recidivism outcomes for both juvenile and adult samples. Research suggests that there may be meaningful differences in the effectiveness of RJ programs across juvenile and adult samples (Suzuki and Wood, 2018). For this reason, data for youth and adult samples were separated and this study focuses on the adult literature only. This study was not pre-registered.

Inclusion criteria. All empirical studies examining the effect of RJ on recidivism or on any other outcome were considered for inclusion using the following criteria: (1) the study was published in English; (2) the sample consisted of adult clients only; (3) each sample was required to be independent from all other samples. If there were overlapping samples, the sample with the longest follow-up period was chosen for inclusion. However, if the smaller sample was subjected to more rigorous analyses or contained less bias, it was chosen for inclusion over the larger sample; (4) the program being evaluated was an RJ program (see next paragraph for more detail); (5) the study must have evaluated the RJ program by utilizing a control group including standard legal procedures (e.g. probation), alternative programming (e.g. non-RJ diversion programs), or receiving no identifiable services (e.g. cautions); and (6) studies must have reported sufficient statistical information on the difference in recidivism and other outcomes across RJ and control groups. If the study results did not include sufficient data for statistical analysis, efforts were made to contact the researchers to retrieve the necessary information.

In determining whether a program was in fact considered RJ, the program had to be defined and identified explicitly as an RJ program in the study description. Since the definition of RJ can also vary, in addition to being described as RJ, the program had to involve some form of reparation of harms and made an explicit effort to include actual

victims and/or members of the community. Programs that only included surrogate victims were not eligible. Victim involvement was defined as a face-to-face between the victim and the client. If a portion of the sample did not include the actual victim or face-to-face interaction but had made an effort to do so, the study was still eligible for inclusion. There were no inclusion or exclusion criteria with respect to sample size or publication date.

For the outcome variables, recidivism included any new offense that resulted in new charges, rearrests, reconvictions, reincarcerations as well as conditional release violations (e.g. suspensions and revocations). Recidivism was further divided into general and violent. General recidivism referred to any new criminal offense regardless of the nature of the crime, while violent recidivism referred to engaging in further criminal acts that were violent in nature (e.g. assault, homicide, manslaughter). Outcomes were categorized as general recidivism if the study did not separate data according to reoffense type (e.g. violent vs general) or referred to the recidivism outcome as "general recidivism." Outcomes were categorized as violent recidivism if the study separated data for violent reoffences or referred to the recidivism outcome as "violent recidivism." A study could contribute effect sizes for both outcomes if the outcomes were clearly defined and enough information was provided to calculate separate effect sizes.

For all other outcomes, the nature of the outcome and how it was defined and measured were noted. The outcome needed to be measured in both the RJ and comparison group to be included. Victim and client satisfaction referred to how satisfied the victims and clients were with their experiences with the RJ process. Procedural justice was defined as victim and client perceptions that they were treated with fairness, dignity, and respect. Client accountability referred to increased feelings of responsibility or remorse or the expression of an apology toward the victim and/or community. Studies reporting on the severity of recidivism most often involved rating the severity level of the reoffense (e.g. using measures such as the Ministry of Justice Seriousness Scale; Spier, 1995) and calculating the mean or median severity for clients across the RJ and comparison groups or indicating the number of clients whose reoffense was more severe than the index offense (e.g. property vs violent offense).

Search terms. A comprehensive search of the literature was conducted in several relevant databases including PsycINFO, Criminal Justice Abstracts, the National Criminal Justice Service, ProQuest Dissertations and Theses, university and government libraries, as well as government websites (e.g. New Zealand Department of Justice, United Kingdom Archived Content) from inception until 30 April 2018. Key terms included restorative justice, victim offender mediation, conference, circle*, family group conference*, mediated dialogue*, recid*, reoffen*, relapse, offen*, and outcome*. See Supplemental Material (S1, p. 1) for more detail on the systematic search strategy. The search in databases, websites, and reference lists was conducted by one reviewer and decisions regarding inclusion were decided between two reviewers.

After duplicate references were removed from the search results, titles, and abstracts of the articles were screened by two independent screeners to determine eligibility for full-text review. Full texts were then evaluated by two independent reviewers for inclusion based on the inclusion criteria. Whenever there were disagreements at this phase,

screeners met to reach a consensus on eligibility. Reference lists of included articles were searched for other relevant studies that were not retrieved from the databases.

Data extraction. A coding form was developed to systematically extract data from the included full texts (see Supplemental Material, S2, p. 2). The major categories of extracted data included study variables (e.g. type of research design, control group type), program variables (e.g. type of RJ program, fidelity/integrity of RJ principles), sample variables (e.g. gender, race, risk level), and outcomes (e.g. recidivism, victim satisfaction). Client risk-level was based on the reported risk-level from actuarial assessment tools (e.g. Offender Group Reconviction Scale; Howard et al., 2009). If actuarial assessments were not used, risk-level was coded based on whether the sample was described as high-risk or low-risk, and if this was unclear based on the information provided or not reported, it was coded as not reported.

Fidelity to RJ principles items were developed and coded based on Zehr and Mika (1998), and included items such as community representation, client requirement to accept responsibility, repairing harm as a primary goal, identifying and addressing victim and client needs, and equal consideration given to the needs and dignity of all parties involved. Given that repairing harm to victims and the community is a cornerstone of RJ (Zehr and Mika, 1998), RJ programs that included community involvement, addressed victim needs, and focused on repairing harm were coded as high adherence to RJ principles, and programs that did not adhere to all three of these principles but adhered to at least two fidelity items were coded as some adherence to RJ principles. RJ programs that only adhered to one or no fidelity items were coded as minimal adherence to RJ programs. Integrity items included adequate staff training in the RJ program, supervisory oversight, and whether the case was closely monitored. RJ programs that fulfilled all three integrity items were coded as good integrity whereas RJ programs that fulfilled one or two integrity items were coded as minimal integrity. Those that did not fulfill any integrity items were coded as poor integrity/major problems.

Risk of bias. Each study was also coded using the Collaborative Outcome Data Committee's Guidelines for the Evaluation of Sexual Offender Treatment Outcome Studies (CODC Guidelines; Beech, 2007; see Supplemental Material, S3, p. 8) as a measure of potential bias. The CODC is a coding guide designed to assess a study's methodological quality and includes 20 items organized into seven categories such as administrative control of the independent variable, experimenter involvement, and sample size. After rating individual items, an overall rating on the judgments of confidence and bias in the study based on the pattern of individual ratings was made to yield a global rating score ranging from 0 (Rejected=low confidence in the results or evidence of considerable bias) to 3 (Strong=high confidence in the study results and minimal evidence of bias).¹ Given that the CODC global rating score was intended to be a form of structured judgment (Beech, 2007), no specific cut off rules were implemented to determine the global rating score. Nevertheless, the CODC global ratings were coded by two independent raters to reduce the possible influence of subjectivity in the coding and to ensure inter-reliability. Any discrepancies in global ratings were discussed and a consensus coding was used.

Interrater reliability. The data were extracted by two independent coders who were graduate students with extensive knowledge of the RJ literature. Interrater reliability procedures were implemented on the larger meta-analysis conducted including both adult and juvenile samples (total studies = 57). An initial sample of 10 studies was randomly selected for training purposes. After reaching a consensus on these 10 studies, the coders began another phase of interrater coding. In the second interrater analysis phase, a new sample of studies were randomly selected in which half were coded without discussion among the coders ($n = 10$) and the other half coded with discussion ($n = 10$) to identify potential coding issues.

The interrater reliability analyses were based on the 10 studies that were independently coded without discussion. The average intraclass correlation coefficient (ICC); (absolute agreement) across continuous variables was .97 (Range = .40–1.00). In addition, the average kappa and percentage agreement across the categorical variables was 84.4% (Range = .80–1.00). Importantly, only 3/83 variables (3.6%) fell below the acceptable standard (i.e. whether the program was mandatory, whether the program was focused on repairing harm, and the level of the structure of the program). After this final round of interrater analysis, decision rules for future coding were implemented for the remaining items that caused discrepancy in the coding. The first author then proceeded to code the remaining studies independently. Given the complexity of the coding, certain decisions were made throughout the coding process as unique issues arose. Questions and answers to all coding queries were documented throughout (see Supplemental Material, S4, p. 10).

Effect size calculations

The effect size used for the analyses pertaining to recidivism was the log odds ratio (LnOR), which compares the experimental and comparison groups on their relative odds of recidivism. For the present analysis, studies with effect sizes below 1 can be interpreted as a positive effect of RJ on recidivism (i.e. reduced recidivism for the RJ group compared to the comparison group) and effect sizes above 1 are interpreted as a negative effect on recidivism (i.e. increased recidivism for the RJ group compared to the comparison group). The following formulas were used to calculate the LnORs and the variance (see Fleiss, 1994). Note that 0.5 is added to each cell in order to include cells with missing data. When LnOR was derived from 2×2 contingency tables, the inverse of the variance (including both within-study and between-study variance) was used in aggregating the findings, giving more weight to studies with larger samples (Borenstein et al., 2009). According to Borenstein et al. (2009), the inverse of the variance is approximately proportional to sample size and minimizes the imprecision of the pooled effect size estimate:

$$\text{LnOR} = \ln[(a + 0.5)(d + 0.5) / ((b + 0.5)(c + 0.5))]$$

$$\text{Variance of LnOR} = 1 / (a + 0.5) + 1 / (b + 0.5) + 1 / (c + 0.5) + 1 / (d + 0.5)$$

For the other outcomes (e.g. victim and client satisfaction), odds ratios were not appropriate as the data were more often presented as mean differences between RJ groups and

comparison groups. Cohen's d with corresponding 95% confidence intervals (CI) was used as the effect size indicator for these outcomes (Hasselblad and Hedges, 1995). When Cohen's d was derived from means and standard deviations (e.g. mean level of victim satisfaction for the RJ group vs the comparison group), the inverse of the variance (including only within-study variance) was used in aggregating the findings, giving more weight to studies with larger samples (Borenstein et al., 2009). If a result was originally reported in a 2×2 contingency table (e.g. client took responsibility for actions (yes/no) divided by RJ and control group), it was converted into Cohen's d , using the following formulas for calculating the d and the variance (see Sánchez-Meca et al., 2003):

$$d = \ln[((a + 0.5)(d + 0.5)) / ((b + 0.5)(c + 0.5))] / 1.65$$

$$\text{Variance of } d = 0.3673(1 / (a + 0.5) + 1 / (b + 0.5) + 1 / (c + 0.5) + 1 / (d + 0.5))$$

The convention for evaluating Cohen's (1988) d interprets values of .20 to be "small," values of .50 to be "medium," and values greater than .80 to be "large."

Aggregating effect sizes. The current meta-analysis calculated both fixed-effect and random-effects models using SPSS Statistical Software Version 27 (both are presented in the tables). With fixed-effect models, all studies are assumed to have the same effect size and the summary effect is an estimation of the common effect size (Borenstein et al., 2009). In contrast, the random-effects model assumes that the true effect size varies across studies. Therefore, the summary effect estimates the mean of the distribution of effect sizes by including an additional between-study error term to capture this variability (Borenstein et al., 2009). Although the random-effects model often provides the more realistic interpretation of the data, it also requires a larger sample size ($k > 30$) in order for the T^2 to properly estimate the between-study variance (Schulze, 2007). With smaller sample sizes, T^2 has poor precision (see Borenstein et al., 2009; Schulze, 2007). Therefore, for some of the outcomes, interpretation of the fixed-effect model may be more appropriate. For this reason, both models are presented in the tables. For recidivism, the random-effects model will be discussed in text given the large sample size whereas for other outcomes, the fixed-effect will be highlighted. Importantly, as between-study variability decreases, the results of both fixed-effect and random-effects models converge.

As a general rule, individual effect sizes can only be meta-analyzed if there are at least three independent effect sizes to include in the analysis (Borenstein et al., 2009). For this reason, outcomes were only included in the analyses if at least three effect sizes were identified across the studies. The number of studies contributing to each meta-analyzed result in the tables in the "Results" section is denoted with a k .

Heterogeneity in the effect size distributions were examined using a Cochran's Q statistic and an I^2 test. The Q statistic assesses whether the overall treatment effects are primarily due to sampling error or to systematic differences among the studies and sampling error (Borenstein et al., 2009). The I^2 test measures the proportion of the observed variance across effect sizes due to treatment effects to unsystematic variance and ranges from 0% to 100% (Higgins et al., 2003). The I^2 values of 25%, 50%, and 75% can be

considered low, moderate, and high variability, respectively (Higgins et al., 2003). Several steps were taken to identify potential outliers in the data adopted from Hanson and Bussière (1998) who considered a study to be an outlier if (1) it is an extreme value, (2) the overall Q is significant, and (3) it accounts for more than 50% of the overall Q . Effect sizes that met these criteria were further examined to identify why there were identified as potential outliers. Analyses were run with and then without the outlier to determine the influence the effect size was having on the results.

The influence of moderator variables was also examined. For categorical variables, Q -between analyses were conducted to test the differences between groups. The Q -between is calculated by conducting a meta-analysis on all studies with a given moderator and partitioning the variability within each level of the moderator (i.e. Q -within), subtracting each level's Q -within value from the total. The following formulas were used to calculate Q -within and Q -between (see Borenstein et al., 2009):

$$Q_{within} = \sum_{j=1}^p Q_j.$$

$$Q_{between} = Q - Q_{within}$$

The Q -between value is then compared to the chi-square distribution with $k-1$ ° of freedom. A significant Q -between statistic suggests that the moderator accounts for a significant proportion of variance from the outcome variable across samples. It should be noted that a moderator could only be analyzed if a minimum of three effect sizes were available per level of the moderator. In addition to the variables that had a sufficient number of effect sizes for moderator analysis, the CODC was used as a moderator on the outcomes to assess for potential risk of bias in the included studies.

Publication bias assessment

To examine the potential influence of publication bias on aggregated findings, the publication status of the articles (i.e. whether they were published in a peer-reviewed journal or not) was analyzed as a potential moderator on the outcomes. We also included a funnel plot to test the presence of publication bias, which assumes that larger studies are more likely to be published while smaller studies are more likely to be missing (Borenstein et al., 2009).

Results

Search results

The initial search resulted in 19,006 records (18,973 from databases, 33 from other sources; see Figure 1). At this point, all abstracts were examined by two independent coders and 18,883 were excluded for not meeting minimum inclusion criteria leaving 173 records for full-text review. Two coders once again went through the full texts and

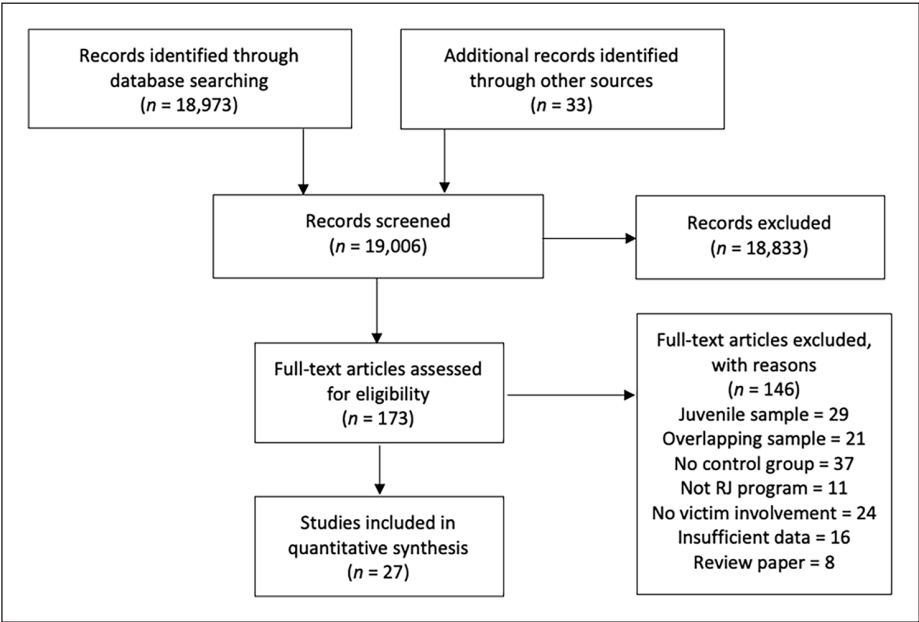


Figure 1. The PRISMA flow diagram for the systematic search.
PRISMA: preferred reporting items for systematic reviews and meta-analyses.
Source: <http://prisma-statement.org>
Other sources include articles retrieved through hand-searching and inter-library loans.

applied all of the inclusion criteria. As can be seen in Figure 1, 146 full texts were excluded leaving a final total of 27 studies representing 34 unique samples for analysis. The full list of studies and samples are in Table 1 (full references for included studies are in the Supplemental Material, S5, pp. 15–18).

Descriptive information

Full tables of descriptive information are provided in the Supplemental Material (S6–S9, pp. 19–26). The majority of studies included predominantly male client samples (i.e. at least two-thirds majority; 88.2%, $k=30$) and for studies that reported the mean age, the overall mean age of the sample was 27.63 ($k=22$). Racial groups were coded according to the racial majority of the sample (i.e. at least two-thirds majority), and if there was no race majority, it was coded as mixed. The racial groups included mostly Caucasian (38.2%, $k=13$), Indigenous (11.8%, $k=4$), Hispanic (2.9%, $k=1$), other minority group (5.9%, $k=2$), and not reported (41.2%, $k=14$). Most of the included samples were assessed as medium risk (58.8%, $k=20$), followed by low risk (17.6%, $k=6$), and high risk (5.9%, $k=2$). The majority of studies were unpublished, non-peer reviewed reports (58.8%, $k=20$) and the evaluators were primarily agency-based researchers (52.9%, $k=18$) or independent non-agency-based researchers (44.1%, $k=15$). When examining

Table 1. List of studies included in the meta-analysis.

Study ID	Authors	Year	Peer reviewed	Country of origin
1.00	Bonta et al.	1983	No	Canada
2.00	Bonta et al.	2002	No	Canada
2.10	Joubert	2004	No	Canada
3.00	Boriboonthana & Sangbuangamlum	2013	No	Thailand
4.00	Davis	2009	No	The United States
5.00	Fitzgerald	2008	No	Australia
6.00	Heinz et al.	1976	Yes	The United States
7.00	Humphrey et al.	2012	No	The United States
8.00	Jones	2009	No	Australia
9.00, 9.01	Maxwell & Morris	2003	Yes	New Zealand
10.00	Meadows et al.	2012	No	The United Kingdom
11.00, 11.01	Miers et al.	2001	No	The United Kingdom
12.00	Mills et al.	2012	Yes	The United States
13.00	Nuffield	1982	No	Canada
14.00	Paulin et al.	2005a	No	New Zealand
15.00	Paulin et al.	2005b	No	New Zealand
16.00	Payne	2017	No	The United States
17.00	Poynton	2013	No	Australia
18.00	Rugge et al.	2005	No	Canada
19.00, 19.01, 19.02, 19.03, 19.04	Shapland et al.	2008	No	The United Kingdom
19.11	Angel et al.	2014	Yes	The United Kingdom
20.00	Sherman et al.	2000	No	Australia
20.10	Barnes	1999	No	Australia
20.20	Barnes et al.	2015	Yes	Australia
21.00	Stewart et al.	2018	Yes	Canada
22.00	Triggs	2005	No	New Zealand
23.00, 23.01	Umbreit et al.	1995	Yes	Canada
24.00, 24.01	Umbreit et al.	1996	No	The United Kingdom
25.00	Walker & Hayashi	2009	Yes	The United States
26.00	Wilcox et al.	2004	No	The United Kingdom

Full references are available in the Supplemental Material (Study ID number = the first decimal place indicates a new article while the second decimal place indicates a unique sample).

the specific source of the unpublished samples, one sample came from an unpublished dissertation while the others were program evaluations completed by government agencies or research centers associated with specific universities.

The most common sampling procedure involved matching RJ participants to a control group (a priori or post hoc equivalency; 47.0%, $k=16$), followed by convenience sampling (i.e. not randomized or matched; 32.4%, $k=11$) and random sampling (20.6%, $k=7$). The average length of follow up was 19.74 months. Methodological quality was

Table 2. Mean odds ratios of the effect of restorative justice for general and violent recidivism.

Recidivism outcome	Mean LnOR	Fixed		Mean LnOR	Random		z	Q	I ²	k	n
		95% CI			95% CI						
		Lower	Upper		Lower	Upper					
General	0.93	0.90	0.97	0.83	0.72	0.95	1.88	238.22***	88.67	28	22,233
Violent	0.87	0.70	1.08	0.88	0.56	1.38	1.00	19.51**	74.37	6	1160

CI: confidence intervals; OR: odds ratio; k: number of samples.
General recidivism encompasses any recidivism outcome. Violent recidivism refers to any violent recidivism outcome. Bolded values represent significant mean effect sizes.
p* < .01; *p* < .001.

based on the rating scheme from the CODC guidelines which included rating individual items and a global score based on overall impressions of methodological quality. When assessing the methodological quality of the studies using the CODC guidelines, the majority were rated as “weak” (35.3%, *k* = 12), followed by “good” (29.2%, *k* = 10) and “rejected” (17.6%, *k* = 6).

Programs were often facilitated by professionals involved in the legal system (35.3%, *k* = 12), followed by volunteers and non-mental health professionals (e.g. trained mediator); each 26.5% (*k* = 9). The RJ programs varied in their degree to which they adhered to principles of RJ. In particular, most involved community representation in the session (52.9%, *k* = 18), face-to-face contact with the victim (88.2%, *k* = 30), required the client to accept responsibility (76.5%, *k* = 26), addressed victim needs (91.2%, *k* = 31) and to a lesser extent, addressed client needs (47.1%, *k* = 16). To achieve a global score of fidelity, an item capturing overall adherence to RJ principles was included and revealed that the majority of programs adhered to some RJ principles (58.8%, *k* = 20), followed by high fidelity to RJ principles (38.2%, *k* = 13) and minimal fidelity to RJ principles (2.9%, *k* = 1). Programs were scored on their overall integrity, where most were revealed to have good integrity (70.6%, *k* = 24), followed by minimal integrity (14.7%, *k* = 5), and poor or major problems (8.8%, *k* = 3). Most RJ programs involved some preparation before the victim and client meetings (67.6%, *k* = 23) with some evidence of internal advocacy/support for the victim or the client (38.2%, *k* = 13).

The effects of RJ on recidivism

The unweighted base rate of recidivism for the RJ group and control group was 32.4% versus 33.8% for general recidivism and 13.5% versus 14.4%, respectively, for violent recidivism. Meta-analyzed results are presented in Table 2. Using the random-effects model (given the large sample size), the mean LnOR of general recidivism was 0.83 (95% CI = [0.73, 0.95]). This translates to a reduction of 17% in the odds of recidivism for the RJ group compared to the control group. There was a significant degree of variability across studies for general recidivism (*Q* = 238.22, *p* < .001), which accounted for

a large proportion of the total between-study variability ($I^2 = 88.67\%$). For violent recidivism, the mean LnOR was 0.88 (95% CI = [0.56, 1.38]), indicating that the odds of recidivism are 12% less likely in the RJ group compared to the control group, which was not significant. Furthermore, there was a significant degree of variability across studies ($Q = 19.51, p < .01$) that accounted for a significant proportion of the total between-study variability ($I^2 = 74.37\%$).

Moderator analyses for general recidivism. The influence of sample, study (Table 3), and program variables (Table 4) was analyzed using the Q -change (Q_d) as the moderator index. In terms of sample moderators, clients who were deemed to be medium-high risk were less likely to reoffend compared to clients who were low-risk ($p < .001$). For study characteristics, studies that used a matched sampling design (e.g. a priori matching or post hoc equivalency) demonstrated greater reductions in recidivism than studies that used a random or convenience/non-equivalence sample ($p < .001$). In addition, studies that were judged to be of poor quality on the CODC guidelines ("Reject") reported greater reductions in recidivism than those that were judged to be weak or good ($p < .001$). This latter finding suggests the possible role of methodological bias, such that lower quality studies are more likely to demonstrate greater program effects than higher quality studies. Finally, for program variables, there were greater reductions in recidivism for programs that did not include community representation (as opposed to those that did; $p < .001$), and for programs that had minimal to some adherence to RJ principles (as opposed to high adherence; $p < .01$). Such findings contradict the RJ literature that posits the importance of adhering to RJ principles in reducing reoffending. Given the small number of studies for violent recidivism, moderator analyses were not possible.

The effects of RJ on other outcomes

Table 5 presents the findings for the effects of RJ on other outcomes. Given the smaller sample size for these analyses, fixed-effects will be discussed; however, results were consistent across both fixed and random-effects analyses. Results indicated that RJ was associated with a moderate significant increase in victim and client satisfaction compared to the control group ($d = .63$ and $.65$, respectively). There was a significant degree of variability across the studies for victim satisfaction ($Q = 16.43, p < .05$), and the variability that was present accounted for more than half of the variability between studies ($I^2 = 69.56\%$). There was also a moderate and significant increase in victims' perceptions of procedural justice ($d = .59$) if they participated in an RJ program compared to control group; perceptions of the clients' procedural justice did not significantly differ between the groups ($d = 0.08$). There was a significant degree of variability across the studies for both victim and client procedural justice accounting for a large portion of the variability ($I^2 = 83.92\%$ and 92.78% , respectively). RJ was also associated with moderate significant increase in client accountability compared to the control group ($d = .52$); however, it is important to note that this analysis included an outlier (study 4.00). Given that the accountability outcome only included three effect sizes, analyses could not be run with and then without the outlier. Conclusions about the effect of RJ on client accountability should therefore be made with caution. Finally, RJ had no significant effect the severity in recidivism ($d = -.08$).

Table 3. Sample and study moderator analyses for general recidivism outcome.

Sample and study moderators	Mean LnOR	95% CI	Q	Q _Δ	I ²	k	n
Average risk level of RJ participants	0.95	0.91	238.22	14.71***	88.67	28	22,233
Low	0.94	0.89	11.41		64.94	6	14,612
Medium to high	0.90	0.84	212.10		89.99	22	6721
Index violent offense for RJ group	0.95	0.91	238.22	0.03	88.67	28	22,233
Yes	0.96	0.81	58.55		93.17	5	2037
No	0.94	0.91	179.64		87.75	23	20,196
Published	0.95	0.91	238.22	1.66	88.67	28	22,233
Yes	0.92	0.86	143.26		93.72	10	6851
No	0.96	0.92	93.30		81.78	18	15,382
Country	0.95	0.91	238.22	0.09	88.67	28	22,233
North America	0.95	0.89	127.44		93.72	9	6302
Other	0.94	0.90	110.69		83.74	19	15,931
Study design	0.95	0.91	238.22	22.16***	88.67	28	22,233
Random	1.11	1.00	34.43		88.38	5	2093
Matched (a priori and post hoc equivalency)	0.90	0.86	114.45		86.02	17	17,630
Convenience/non-equivalence	1.15	1.01	67.18		92.56	6	2510
Control group composition	0.95	0.91	238.22	1.52	88.67	28	22,233
Eligible, not referred	0.93	0.89	184.60		89.17	21	15,981
Refused, non-eligible, or mixed/unclear	0.98	0.92	52.10		88.48	7	6252
Recidivism for RJ based on?	0.95	0.91	238.22	3.46	88.67	28	22,233
Admissions	1.03	0.93	83.03		93.98	6	3952
Graduates	0.93	0.90	151.73		86.16	22	18,281
CODC outcome	0.95	0.91	238.22	28.72***	88.67	28	22,233
Reject	0.66	0.58	26.11		77.02	7	2302
Weak	0.99	0.94	151.87		90.78	15	9810
Good	0.95	0.89	31.52		84.14	6	10,121

CI: confidence interval; RJ: restorative justice; OR: odds ratio; k: number of samples; Q_Δ: Q-change; CODC: collaborative outcome data committee guidelines. Bolded values represent significant mean effect sizes. ***: p < .001.

Table 4. Program moderator analyses for general recidivism outcome.

Program moderators	Mean LnOR	95% CI	Q	Q _A	I^2	k	n
Type of RJ program	0.95	0.91	238.22	0.86	88.67	28	22,233
Conference	0.95	0.91	121.12		84.31	20	20,651
Mediation, circles, or restitution program	0.88	0.76	116.24		93.98	8	1582
Stage of criminal justice process	0.95	0.91	238.22	0.02	88.67	28	22,233
Pre-arrest/pre-charge, post-charge	0.94	0.90	160.91		85.71	24	16,819
Post-sentencing, but pre-release	0.95	0.89	77.29		96.12	4	5414
Community representation	0.95	0.91	238.22	17.01***	88.67	28	22,233
Yes	1.01	0.96	118.23		86.47	17	11,476
No	0.86	0.81	102.98		90.29	11	10,757
Client required to accept responsibility	0.95	0.91	238.22	3.45	88.67	28	22,233
Yes	0.95	0.92	178.72		88.25	22	20,719
No	0.79	0.65	56.05		91.08	6	1514
Attempts to identify and address client needs	0.95	0.91	238.22	1.17	88.67	28	22,233
Yes	0.91	0.84	137.57		90.55	14	4821
No	0.96	0.92	99.48		86.93	14	17,412
Attempts to identify and address victim needs	0.95	0.91	238.22	0.01	88.67	28	22,233
Yes	0.95	0.91	226.10		89.39	25	20,556
No	0.94	0.77	12.11		83.49	3	1677
RJ model and structure	0.95	0.91	238.22	0.52	88.67	28	22,233
Vague	0.99	0.87	67.24		86.62	10	3468
Moderate or detailed	0.94	0.90	170.46		90.03	18	18,765
Who administers RJ?	0.95	0.91	238.22	0.04	88.67	28	22,233
Legal professional	0.95	0.90	59.94		83.32	11	12,314
Volunteers	0.95	0.89	80.65		92.56	7	6114
Other (e.g. mental health professional)	0.94	0.84	97.59		90.78	10	3807
Adherence to RJ principles	0.95	0.91	238.22	8.65**	88.67	28	22,233
Minimal to some	0.92	0.88	144.38		89.61	16	16,991
High	1.04	0.97	85.19		87.09	12	5242
Integrity of RJ services	0.95	0.91	238.22	0.78	88.67	28	22,233
Good/very good	0.95	0.91	189.31		89.44	21	19,624
Poor/major problems	0.89	0.77	48.13		87.53	7	2609

CI: confidence interval; RJ: restorative justice; k: number of samples; Q_A: Q-change.

Bolded values represent significant mean effect sizes.

** $p < .01$; *** $p < .001$.

Table 5. Mean Cohen's *d* of the effect of restorative justice on other outcomes.

Outcome	Mean Cohen's <i>d</i>		Mean Cohen's <i>d</i>		<i>z</i>	<i>Q</i>	<i>I</i> ²	<i>k</i>	<i>n</i>
	Fixed	95% CI	Random						
			Lower	Upper					
Victim satisfaction	0.63	0.45 0.80	0.39 0.88	0.64	7.00	10.03	40.16	7	1087
Client satisfaction	0.65	0.50 0.80	0.50 0.80	0.65	8.13	0.13	0.00	5	716
Victim procedural justice	0.59	0.46 0.72	0.30 1.08	0.69	8.43	31.10***	83.92	6	1003
Client procedural justice	0.08	-0.07 0.23	-0.17 1.40	0.62	1.00	55.39***	92.78	5	716
Client accountability	0.52	0.35 0.70	0.35 0.70	0.52	5.78	1.81	0.00	3	531
Severity of recidivism	-0.08	-0.21 0.04	-0.22 0.05	-0.09	1.14	4.22	5.25	5	948

CI: confidence interval; *k*: number of samples.

Victim and client outcomes were indexed using self-report data from participants; severity of recidivism was measured based on type of recidivism offense where higher values indicated higher severity; bolded values represent significant mean effect sizes.

****p* < .001.

Moderator analyses for other outcomes. Given the small sample size, only the influence of whether the program was mandatory (yes vs no) and the overall integrity of the program (good vs minimal integrity) could be examined as potential moderators for perceptions of victim procedural justice (full table is available in the Supplemental Material, S10, p. 27). Only the level of integrity of the RJ program had a significant moderating effect on victim procedural justice ($p < .001$). When compared to the control group, RJ programs that had good integrity (vs those with minimal integrity) had significantly higher ratings of victim procedural justice ($d = .84$ vs $.38$).

Publication bias

The majority of the studies included in the analyses were unpublished (58.8%, $k=20$) and publication status was not a significant moderator of the effect of RJ on general recidivism (publication status could not be examined for other outcomes given the small sample size). In order to provide a detailed visualization of the potential effects of publication bias, funnel plots were generated with the effect sizes on the X -axis and the standard errors and sample sizes on the Y -axis. When examining the funnel plots, the studies appeared to be distributed symmetrically about the mean effect size, suggesting a lack of publication bias (figures can be found in the Supplemental Material, S11–12, p. 28).

Discussion

Overall, our results suggest that RJ programs led to a significant and small reduction in general recidivism compared to traditional legal system procedures and approaches, and no significant effect on violent recidivism. Furthermore, RJ only demonstrated significant reductions in general recidivism when the study quality was rated as poor (i.e. rejected) based on the CODC guidelines; when studies were rated as weak or good, RJ had no significant effect on general recidivism. RJ programs were also related to significant and moderate increases in victim and client satisfaction, victims' perceptions of procedural justice and fairness, and overall client accountability.

The effects of RJ on recidivism

A considerable amount of work has been devoted to elucidating the promise of RJ programs as a more effective and humane alternative to traditional legal system procedures. Through its focus on repairing harm, client accountability and reintegration, RJ is purported to eventually lead to reductions in recidivism (Zehr, 2002). However, consolidating quantitative studies comparing RJ's effectiveness to traditional legal system procedures over the past 40 years revealed only modest reductions in general recidivism. Of course, once study quality is considered, RJ's effect on reduced recidivism may actually reflect methodological choices on the part of researchers as opposed to actual reductions in recidivism. After accounting for study quality through the CODC guidelines, RJ's negligible effect on general recidivism is more reflective of theoretical work suggesting that RJ programs are designed to emphasize reparation rather than reductions in recidivism; therefore, RJ programs may not be as effective as programs that make

reducing recidivism the main focus such as those falling under the RNR theoretical framework (Walgrave et al., 2021; Zehr, 2002). These findings suggest that incorporating RNR principles may increase the effectiveness of RJ programs in reducing recidivism.

RJ was unrelated to reductions in violent recidivism compared to traditional legal system procedures. However, these results should be considered with caution given that there were only a small number of studies reporting the effects of RJ on violent recidivism. RJ programs are not as frequently used for violent index offenses compared to clients with non-violent index offenses (e.g. property crimes, driving-related crimes; Brooks, 2013). In accordance with the RNR model, RJ may not be as effective for individuals who are higher risk given that the intensity of the program may not match the severity of the crime or the client's needs (Bonta et al., 2006). In addition, given its focus on victim needs, RJ programs are designed to place less emphasis on rehabilitation for offending populations which are often needed among higher risk groups (Andrews and Bonta, 2010). In one previous systematic review, while the results indicated that RJ may be more effective for individuals with violent offenses, the comparison was made specifically between violent versus property offenses, and the overall comparison was not statistically significant (Strang et al., 2013).

When taking a more detailed look at the moderators for the relationship between RJ and general recidivism, medium- to high-risk clients demonstrated greater program effects than low-risk clients when compared to the control group. This may be due to a floor effect; clients who are lower risk have less room for movement compared to clients in the medium and higher risk groups. Furthermore, only a small number of studies reported the use of standardized risk assessment tools to evaluate risk-level, and hence, the quality of risk classification in the majority of the studies was actually unknown.

Two program features also moderated the relationship between RJ programs and general recidivism: community representation and adherence to RJ principles. RJ programs that did *not* include community representation demonstrated *greater* reductions in recidivism than programs that did, which is inconsistent with literature on the importance of community representation on desistance (Dzur and Olson, 2004). However, this result was also found in a previous meta-analysis (Bain, 2012). One possible explanation is that the presence of community members not directly tied to the criminal act may counteract the reparation process if there is hostility directed toward the client (Cook, 2006). In addition, it is possible that the process of reintegrative shaming involving community members may have an aversive impact on some clients, leading to increases in recidivism (Bain, 2012). Given these unexpected results, future research should further examine the benefits of community involvement in the RJ process and whether the perceived benefits outweigh the potential negative impact on reoffending.

Most notably, RJ programs with minimal to some adherence to RJ principles demonstrated greater program effects than programs with high adherence. According to the literature, adherence to RJ principles (e.g. reparation, attending to victim needs) is key for optimizing RJ program outcomes (McCold, 2000). However, these results suggest that programs that had a greater alliance to RJ principles were less beneficial for the purpose of reducing recidivism. These results may indicate that "purist" RJ programs are less interested in client needs than victim and community needs and therefore place less

emphasis on reducing recidivism. A previously mentioned, RJ theory does not align closely with theories of offender rehabilitation designed to reduce recidivism such as RNR theory. Reducing recidivism is often not the key goal of RJ, but rather a potential by product of the process. However, it is also possible that the difficulty with defining and measuring adherence to RJ principles could have influenced these results. While a detailed rating system was used to assess overall adherence, many studies did not provide the necessary information to score the item. Given these findings, more research is needed to articulate which RJ principles may conflict with the goal of reducing recidivism to maximize RJ program outcomes.

The moderator analyses also revealed unexpected findings. Studies including matched sample designs demonstrated greater reductions in general recidivism compared to convenience sampling and random sampling designs. Given that studies with lower methodological quality result in greater program effects, it would be expected that convenience sampling would be associated with greater reductions in recidivism compared to matched designs. One possibility for these discrepant findings may be the greater number of studies using matched designs compared to those using randomized designs or convenience sampling. The unbalanced distribution of studies across different sampling designs could have resulted in insufficient power for the moderator analysis (Cuijpers et al., 2021), further highlighting the need for using randomized designs in future research.

Importantly, other moderators did not reach significance, including whether the client was required to accept responsibility, whether the program addressed victim and client needs, type of RJ program, and the extent to which the RJ program model had a clear and detailed structure. The first two align with the finding that adherence to RJ principles (e.g. client accountability, addressing victim needs) may not be as important for reducing recidivism as RJ theory suggests. For example, while holding the client accountable for their criminal act may be beneficial for victims and the community and promote client reintegration, accountability is not a risk factor that is related to reoffending as opposed to validated criminogenic needs (e.g. substance use, procriminal attitudes; Andrews et al., 2011). In addition, while addressing victim and client needs may be important for improving program satisfaction, it may be unrelated to client recidivism risk (e.g. victims feeling safe and supported).

Finally, while the latter two findings suggest that the type of RJ program and the extent to which the program was structured was not associated with significant reductions in recidivism, these results may be due to imprecision with defining subgroups for moderator analysis. For example, conferences were compared to three different types of RJ programs in one group (i.e. mediation, circles, restitution programs), while structured programs were compared to programs that had either a low structure or vague structure. Subgroups were collapsed in this way due to the lack of a sufficient sample size to separate all possible categories. More research with larger sample sizes is needed to determine whether these moderators would reach significance with more clearly defined subgroups.

The effects of RJ on other outcomes

While RJ may not be as effective for reducing recidivism as the theoretical literature suggests, it is important to examine its influence on other outcomes. Here, RJ clearly demonstrated advantages over traditional legal system procedures. RJ was associated with

increases in client and victim satisfaction. For victims, such results were expected given that RJ is designed to address the needs of victims who have historically been neglected in traditional legal systems (Woolford, 2009). Improvements in client satisfaction were also expected given the emphasis on punishment in traditional approaches, which contrasts RJ's humanistic, person-centered approach (Strang et al., 1999). Victim satisfaction was significantly moderated by the integrity of the RJ program such that higher integrity was associated with greater program effects. It is possible that programs with greater integrity (e.g. properly trained staff, supervisory oversight) make it easier for victims to partake in the difficult process of confronting the client (Stubbs, 2007), and may also simply be better organized creating less stress for participants.

RJ was also associated with greater increases in victims' perceptions of procedural justice compared to traditional procedures. Considering the neglect of victim needs in the legal system, as well as the emphasis on victim empowerment in RJ, it is not surprising that victims would experience a greater level of procedural justice in RJ programs. Furthermore, victim procedural justice was also found to be significantly moderated by the integrity of the RJ program, such that higher integrity was associated with greater program effects. This demonstrates the importance of RJ programs to include good structured guidelines, high-quality staff training, and appropriate oversight if positive outcomes are to be optimized.

The results also provided preliminary evidence that RJ increases client accountability over traditional approaches. The results of this outcome were largely based on a mix of self-report assessments of accountability for the criminal act, victim perceptions of client accountability, and whether or not the client expressed feelings of accountability, such as through apologizing. Given that encouraging feelings of remorse and guilt is often built into the RJ process (Hayes and Daly, 2003), it would be expected that RJ enhances client accountability over traditional legal system processes. However, while clients may have expressed their accountability, it cannot be determined whether clients experienced genuine responsibility for their actions. Future research should further examine whether RJ improves client accountability using valid self-report measures.

Our results suggest that RJ was not associated with greater decreases in the severity of recidivism compared to traditional legal system procedures, contrary to some past research (e.g. Urban and Burge, 2006). Aside from the claim that RJ programs reduce recidivism (e.g. Zehr, 2002), there is little discussion about why RJ programs are expected to reduce the severity of recidivism. It should also be noted that there were only five effect sizes included in the analysis of offense severity. More research is needed to determine whether RJ has advantages over traditional legal system procedures in terms of its ability to reduce the severity of recidivism.

Limitations

While the current meta-analysis has many strengths (e.g. consideration of a number of important moderators; more objective measure of study quality), several limitations must be noted. The results of meta-analyses are only as accurate and reliable as the included articles. The influence of articles that were not found, despite the broad search parameters, is still possible. While our search strategy included several databases and adhered to systematic review guidelines (Page et al., 2021; Siddaway et al., 2019), it is possible that

eligible articles were missed since not all possible databases were searched. The same applies for the gray literature search. It is worth noting however that the number of gray literature sources that were searched may be less of a concern given that there was little evidence of publication bias in this study. Our search was also limited to articles written in English, which reduces generalizability. Another limitation is the stringent criteria for including studies in the analyses. While inclusion rules could be viewed as a strength of this study by only considering optimal forms of RJ, these same rules exclude a wider range of studies examining RJ practices, particularly those not involving direct victims. The results of the current meta-analysis must therefore be interpreted within the context of the existing search strategy and inclusion criteria.

Furthermore, as previously articulated, many of the items scored by the coders (e.g. adherence to RJ principles, integrity of the RJ program) were vulnerable to some degree of subjectivity due to variations in definitions in RJ theory as well as the limited information provided in the articles. Relatedly, resource constraints limited our ability to double-code all articles which is often recommended for extracting data for meta-analyses (Page et al., 2021); although through multiple rounds of interrater analyses, more than 50% of articles were double-coded. The study methodology was also not pre-registered despite this being a recommended practice for meta-analytic research, which introduces the possible influence of bias in the results (Lakens et al., 2016). Despite these coding-related issues, our interrater procedure provided substantial protection against subjectivity and other potential biases in the coding. A related issue was the scoring for program adherence to RJ principles and integrity, which revealed the opposite findings than what was expected. While the interrater reliability for these items reached acceptable levels, these items may not adequately capture program fidelity and integrity. Thus, other methods for measuring these variables should be explored in future research.

Recommendations for policy and research

Considering the above limitations, there are several recommendations that can still be made to potentially improve RJ research and practice. It may be worth considering integrating RJ programs with more empirically supported models of offender rehabilitation in order to maximize reductions in recidivism. Of course, this assumes that the goal of the RJ program is in fact to reduce recidivism. It seems that RJ programs, on their own, are unlikely to have perceptible impacts on recidivism outcomes. By combining different approaches, reductions in recidivism could be optimized while providing a greater role for victims within the more traditional RNR programs.

A major insight from the results of this study was the effectiveness of RJ programs in enhancing other outcomes, namely, victim and client satisfaction, victim procedural justice, and client accountability. These results suggest that RJ programs have the unique advantage of improving outcomes for victims who are often neglected in the legal system. The implementation of RJ programs should therefore consider enhancing other outcomes for both victim and clients as a main priority. Future research is needed on how RJ programs can tailor their practices to further maximize these other outcomes. However, clearly, higher program integrity, including appropriate structure and staff training and good program oversight appear key to having successful outcomes for victims involved in RJ.

The most troubling finding was the lack of high-quality research evaluating the effectiveness of RJ programs in reducing recidivism and improving other outcomes. Methodological improvements that are needed include using random sampling methods, measuring outcomes based on admissions rather than graduates, and adequately controlling for group differences. Many of the studies did not explore group differences on basic characteristics such as gender and age and often did not report attrition rates or loss of data. Thus, these results suggest that there are still major methodological improvements to be made in RJ research.

Conclusion

In summary, the results of the present analysis of studies comparing the effectiveness of RJ to traditional legal system responses demonstrated that RJ was related to small reductions in general recidivism. The purported effectiveness of RJ programs in reducing recidivism may be driven in part by the influence of study methodological quality in which lower quality studies indicate greater program effects than higher quality studies. It is also important to consider this study's limitations such as the possible exclusion of eligible articles and subjectivity in coding procedures. Nevertheless, RJ showed promise in enhancing other victim and client outcomes such as satisfaction, procedural justice and fairness, and client accountability, which represent important advantages of RJ programs over traditional legal responses. Taken together, while such findings underscore the valuable role of RJ programs in the legal system, further research is needed to investigate how RJ could be integrated with other programs that target recidivism more effectively.

Acknowledgements

The authors would like to thank Public Safety Canada for initiating this project and overseeing earlier aspects of the project. They are also grateful for Nicholas Chadwick and Kayla A. Wanamaker for updating the initial search and conducting a review of potential articles for inclusion. Finally, they are grateful for the meta-analysis syntax created and shared by Maaïke L. Helmus and Kelly M. Babchishin.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Julie Blais  <https://orcid.org/0000-0001-7386-9549>

Supplemental Material

Supplemental material for this article is available online.

Note

1. Note that the “Rejected” category represented studies with the lowest methodological quality on the Collaborative Outcome Data Committee (CODC). These studies were not rejected from inclusion in the current analyses.

References

- Andrews DA and Bonta J (2010) *The Psychology of Criminal Conduct*, 5th edn. New Providence, NJ: LexisNexis Matthew Bender.
- Andrews DA, Bonta J and Wormith SJ (2011) The Risk-Need-Responsivity (RNR) model: Does adding the good lives model contribute to effective crime prevention? *Criminal Justice and Behavior* 38(7): 735–755.
- Andrews DA, Zinger I, Hoge R, et al. (1990) Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology* 28: 369–404.
- Bain K (2012) *Restorative justice and recidivism: A meta-analysis*. Master’s Thesis, ProQuest Dissertations & Theses Global (Order No. 1532051), Ann Arbor, MI.
- Bazemore G and Schiff M (2001) *Restorative Community Justice: Repairing Harm and Transforming Communities*. New York: Routledge.
- Bazemore G and Umbreit M (1999) Conferences, circles, boards, and mediations: Restorative justice and citizen involvement in the response to youth crime (Funded by Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice) (Prepared for the Balanced and Restorative Justice Project). Available at: <https://www.ojp.gov/ncjrs/virtual-library/abstracts/conferences-circles-boards-and-mediations-restorative-justice-and>
- Beech A (2007) *The Collaborative Outcome Data Committee’s Guidelines for the Evaluation of Sexual Offender Treatment Outcome Research: Part 2: CODC Guidelines*. Ottawa, ON, Canada: Public Safety Canada.
- Bonta J, Bourgon G, Rugge T, et al. (2010) *The Strategic Training Initiative in Community Supervision: Risk-Need-Responsivity in the Real World* (User Report 2010-01). Ottawa, ON, Canada: Public Safety and Emergency Preparedness Canada.
- Bonta J, Jessemann R, Rugge T, et al. (2006) Restorative justice and recidivism: Promises made, promises kept? In: Sullivan D and Tift L (eds) *Handbook of Restorative Justice: A Global Perspective*. New York: Routledge, pp. 108–120.
- Borenstein M, Hedges L, Higgins J, et al. (2009) *Introduction to Meta-Analysis*. Chichester: John Wiley & Sons.
- Bradshaw W, Roseborough D and Umbreit MS (2006) The effect of victim offender mediation on juvenile offender recidivism: A meta-analysis. *Conflict Resolution Quarterly* 24(1): 87–98.
- Braithwaite J (1999) Restorative justice: Assessing optimistic and pessimistic accounts. In: Tonry M (eds) *Crime and Justice: A Review of Research*. Chicago, IL: University of Chicago Press, pp. 1–127.
- Brooks A (2013) *Moving forward: Two approaches to repairing the harm through restorative justice*. Doctoral Dissertation. ProQuest Dissertations & Theses Global (Order No. 3597003), Ann Arbor, MI.
- Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.
- Cook K (2006) Doing difference and accountability in restorative justice conferences. *Theoretical Criminology* 10(1): 107–124.
- Cuijpers P, Griffin JW and Furukawa TA (2021) The lack of statistical power of subgroup analyses in meta-analyses: A cautionary note. *Epidemiology and Psychiatric Sciences* 30: e78.

- Daly K (2001) Conferencing in Australia and New Zealand: Variations, research findings, and prospects. In: Morris A and Maxwell G (eds) *Restorative Justice for Juveniles: Conferencing, Mediation and Circles*. Oxford: Hart Publishing, pp. 59–84.
- Department of Justice Canada (2018) The effects of restorative justice programming: A review of the empirical (RR2000-16e). *Department of Justice Canada*. Available at: https://www.justice.gc.ca/eng/rp-pr/csj-sjc/jsp-sjp/rr00_16/toc-tdm.html
- Dzur AW and Olson SM (2004) The value of community participation in restorative justice. *Journal of Social Philosophy* 35(1): 91–107.
- Federal-Provincial-Territorial Working Group on Restorative Justice (2020) Increasing the use of restorative justice in criminal matters in Canada: Baseline report. Available at: <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/2020-resjus-jusrep/2020-resjus-jusrep-en.pdf>
- Fleiss JL (1994) Measures of effect size for categorical data. In: Cooper H and Hedges LV (eds) *The Handbook of Research Synthesis*. New York: Russell Sage, pp. 245–260.
- Gang D, Loff B, Naylor B, et al. (2021) A call for evaluation of restorative justice programs. *Trauma, Violence, & Abuse* 22(1): 186–190.
- Gavrielides T and Worth P (2014) Another push for restorative justice: Positive psychology and offender rehabilitation. In: Pearson MH (ed.) *Crime: International Perspectives, Socioeconomic Factors and Psychological Implications*. Hauppauge, NJ: Nova Science Publishers, pp. 161–182.
- Hanson RK and Bussière MT (1998) Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology* 66(2): 348–362.
- Hasselblad V and Hedges LV (1995) Meta-analysis of screening and diagnostic tests. *Psychological Bulletin* 117: 167–178.
- Hayes H and Daly K (2003) Youth justice conferencing and reoffending. *Justice Quarterly* 20: 725–764.
- Higgins JP, Thompson SG, Deeks JJ, et al. (2003) Measuring inconsistency in meta-analyses. *BMJ* 327: 557–560.
- Howard P, Francis B, Soothill K, et al. (2009) OGRS3: The revised offender group reconviction score. *Research Summary* 7: 1–6.
- Lakens D, Hilgard J and Staaks J (2016) On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychology* 4(1): 24.
- Latimer J, Dowden C and Muise D (2005) The effectiveness of restorative justice practices: A meta-analysis. *The Prison Journal* 85(2): 127–144.
- Looman J and Abracen J (2013) The risk need responsivity model of offender rehabilitation: Is there really a need for a paradigm shift? *International Journal of Behavioral Consultation and Therapy* 8(3–4): 30–36.
- McCold P (2000) Toward a holistic vision of restorative juvenile justice: A reply to the maximalist model. *Contemporary Justice Review* 3: 357–414.
- Marshall TF (1996) The evolution of restorative justice in Britain. *European Journal on Criminal Policy and Research* 4: 21–42.
- Maxwell G and Morris A (2003) Putting restorative justice into practice for adult offenders. *The Howard Journal* 40(1): 55–69.
- McCold P and Wachtel B (2012) *Restorative Policing Experiment: The Bethlehem Pennsylvania Police Family Group Conferencing Project* (Reprinted from 1998). Eugene, OR: Wipf and Stock Publishers.
- McGarrell EF and Hipple NK (2007) Family group conferencing and reoffending among first-time juvenile offenders: The Indianapolis experiment. *Justice Quarterly* 24(2): 221–246.
- Page MJ, McKenzie JE, Bossuyt PM, et al. (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372: n71.

- Piggott E and Wood WR (2018) Does restorative justice reduce recidivism? Assessing evidence and claims about restorative justice and conferencing. In: Gavrielides T (ed.) *International Handbook of Restorative Justice*. New York: Routledge, pp. 1–22.
- Sánchez-Meca J, Marín-Martínez F and Chacón-Moscoso S (2003) Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods* 8(4): 448–467.
- Schulze R (2007) The state and the art of meta-analysis. *Zeitschrift für Psychologie / Journal of Psychology* 215(2): 87–89.
- Shapland J, Atkinson A, Atkinson H, et al. (2008) *Does Restorative Justice Affect Reconviction? The Fourth Report from the Evaluation of Three Schemes*. London: Ministry of Justice.
- Sherman LW, Strang H, Mayo-Wilson E, et al. (2015) Are restorative justice conferences effective in reducing repeat offending? Findings from a Campbell systematic review. *Journal of Quantitative Criminology* 31: 1–24.
- Siddaway AP, Wood AM and Hedges LV (2019) How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology* 70: 747–770.
- Spier P (1995) *Conviction and Sentencing of Offenders in New Zealand: 1985 to 1994*. Wellington, New Zealand: Strategic Assessment, Ministry of Justice.
- Strang H, Barnes G, Braithwaite J, et al. (1999) *Experiments in Restorative Policing: A Progress Report on the Canberra Reintegrative Shaming Experiments (RISE)*. Canberra, ACT, Australia: Australian National University. Available at: <https://www.ojp.gov/ncjrs/virtual-library/abstracts/experiments-restorative-policing-progress-report-canberra>
- Strang H, Sherman LW, Mayo-Wilson E, et al. (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 Systematic Reviews* 9(1): 1–59.
- Stubbs J (2007) Beyond apology? Domestic violence and critical questions for restorative justice. *Criminology & Criminal Justice* 7(2): 169–187.
- Suzuki M and Wood WR (2018) Is restorative justice conferencing appropriate for youth offenders? *Criminology & Criminal Justice* 18(4): 450–467.
- Tyler TR (1990) *Why People Obey the Law: Procedural Justice, Legitimacy, and Compliance*. New Haven, CT: Yale University Press.
- Umbreit MS (1998) Restorative justice through victim-offender mediation: A multi-site assessment. *Western Criminology* 1(1): 1–28. Available at: <http://www.westerncriminology.org/documents/WCR/v01n1/Umbreit/umbreit.html>
- Umbreit MS (2001) *The Handbook of Victim Offender Mediation: An Essential Guide for Practice and Research*. San Francisco, CA: Jossey-Bass.
- Umbreit MS and Coates RB (1992) *Victim Offender Mediation: An Analysis of Programs in Four States of the U.S.* Minneapolis, MN: Minneapolis Citizens Council on Crime and Justice.
- Umbreit MS, Coates RB and Vos B (2004) Victim-offender mediation: Three decades of practice and research. *Conflict Resolution Quarterly* 22: 279–303.
- United Nations Office on Drugs and Crime (2020) Handbook on restorative justice programmes: Second edition. https://www.unodc.org/documents/justice-and-prison-reform/20-01146_Handbook_on_Restorative_Justice_Programmes.pdf
- Urban L and Burge S (2006) Victim/offender mediation in St. Louis: An assessment. *Justice Research and Policy* 8(2): 89–113.
- Walgrave L, Ward T and Zinsstag E (2021) When restorative justice meets the good lives model: Contributing to a criminology of trust. *European Journal of Criminology* 18(3): 444–460.
- Ward T, Fox KJ and Garber M (2014) Restorative justice, offender rehabilitation and desistance. *Restorative Justice* 2(1): 24–42.

- Wong JS, Bouchard J, Grave J, et al. (2016) Can at-risk youth be diverted from crime? A meta-analysis of restorative diversion programs. *Criminal Justice and Behavior* 43(10): 1310–1329.
- Woolford A (2009) *The Politics of Restorative Justice: A Critical Introduction*. Black Point, NS, Canada: Fernwood.
- Zehr H (2002) *The Little Book of Restorative Justice*. Old Philadelphia Pike, PA: Good Books.
- Zehr H and Mika H (1998) Fundamental concepts of restorative justice. *Contemporary Justice Review* 1: 47–55.
- Zernova M (2009) Integrating the restorative and rehabilitative models: Lessons from one family group conferencing project. *Contemporary Justice Review* 12(1): 59–75.

Author biographies

Lindsay Fulham (MA) is a clinical doctoral student in the Department of Psychology at York University. Her research explores the interpersonal factors that cause and maintain borderline personality disorder symptoms as well as effective strategies for treating these symptoms.

Julie Blais (PhD) is an Associate Professor in the Department of Psychology and Neuroscience at Dalhousie University. She is the director of the Forensic Psychology and Personality Lab and conducts research on topics in risk assessment, risk communication, and personality.

Tanya Rugge (BA Law, PhD Psychology) has been conducting research in the area of corrections and criminal justice, including evaluating several restorative justice programs, for over two decades.

Elizabeth A Schultheis (MA) is a doctoral candidate at Carleton University in experimental psychology. She conducts research on matters relating to policing, risk assessment, and intimate partner violence.