

Twelve experiments in restorative justice: the Jerry Lee program of randomized trials of restorative justice conferences

Lawrence W. Sherman^{1,2} · Heather Strang¹ ·
Geoffrey Barnes³ · Daniel J. Woods⁴ ·
Sarah Bennett⁵ · Nova Inkpen⁶ ·
Dorothy Newbury-Birch⁷ · Meredith Rossner⁸ ·
Caroline Angel¹ · Malcolm Mearns⁹ ·
Molly Slothower^{10,11}

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract

Objectives We conducted and measured outcomes from the Jerry Lee Program of 12 randomized trials over two decades in Australia and the United Kingdom (UK), testing an identical method of restorative justice taught by the same trainers to hundreds of police officers and others who delivered it to 2231 offenders and 1179 victims in 1995–

✉ Lawrence W. Sherman
ls434@cam.ac.uk

¹ Jerry Lee Centre of Experimental Criminology, Institute of Criminology, University of Cambridge, Cambridge, UK

² Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA

³ Jerry Lee Center of Criminology, University of Pennsylvania, Philadelphia, PA, USA

⁴ Police Foundation, Washington, DC, USA

⁵ University of Queensland, St Lucia, Australia

⁶ Australian Capital Territory Government, Canberra, Australia

⁷ School of Health and Social Care, Teesside University, Middlesbrough, UK

⁸ London School of Economics, London, UK

⁹ DATACOL, Canberra, Australia

¹⁰ Cambridge Centre for Evidence-Based Policing, College Park, MD, USA

¹¹ University of Maryland, College Park, MD, USA

2004. The article provides a review of the scientific progress and policy effects of the program, as described in 75 publications and papers arising from it, including previously unpublished results of our ongoing analyses.

Methods After random assignment in four Australian tests *diverting* criminal or juvenile cases from prosecution to restorative justice conferences (RJC), and eight UK tests of *supplementing* criminal or juvenile proceedings with RJC, we followed intention-to-treat group differences between offenders for up to 18 years, and for victims up to 10 years.

Results We distil and modify prior research reports into 18 updated evidence-based conclusions about the effects of RJC on both victims and offenders. Initial reductions in repeat offending among offenders assigned to RJC (compared to controls) were found in 10 of our 12 tests. Nine of the ten successes were for crimes with personal victims who participated in the RJC, with clear benefits in both short- and long-term measures, including less prevalence of post-traumatic stress symptoms. Moderator effects across and within experiments showed that *RJCs work best for the most frequent and serious offenders* for repeat offending outcomes, with other clear moderator effects for poly-drug use and offense seriousness.

Conclusions RJ conferences organized and led (most often) by specially-trained police produced substantial short-term, and some long-term, benefits for both crime victims and their offenders, across a range of offense types and stages of the criminal justice processes on two continents, but with important moderator effects. These conclusions are made possible by testing a new kind of justice on a *programmatically* basis that would allow prospective meta-analysis, rather than doing one experiment at a time. This finding provides evidence that funding agencies could get far more evidence for the same cost from programs of identical, but multiple, RCTs of the identical innovative methods, rather than funding one RCT at a time.

Keywords Restorative justice · Randomized controlled trials · Policing · Prospective meta-analysis · Recidivism · Race · Aboriginal Australians · Personal victim offenses · Procedural justice · Moderator effects · Crime harm index

Summary of key findings

Effects of RJC on victims Victims randomly assigned to attend restorative justice conferences (RJC) with their offenders were less fearful of repeat attack by the same person, more pleased with the way their case was handled, and less desirous of violent revenge against their offenders, after receiving far more offender apologies and satisfaction with their justice than control victims. London robbery and burglary victims assigned to RJC, especially females, suffered much less post-traumatic stress than controls, while Canberra victims of violent and property crimes had less emotional impact from the crime than controls for at least 10 years after the arrest of their offenders.

Main effects of RJC on offenders The average effect of RJC on offenders in both Australia and the United Kingdom (UK) is to reduce the frequency of repeat offending

after 2 years, with high cost-effectiveness in all UK tests. While we have no long-term evidence on recidivism in the UK, Canberra evidence to date shows no main effects on recidivism after 15 or more years.

Moderator effects of RJC on recidivism Relative to control groups, RJCs generally failed to reduce recidivism for property crimes, but consistently did so for violent crime. RJCs had the biggest effects on reducing recidivism with high-frequency offenders, but was ineffective or criminogenic for those offenders with medium rates of offending. RJCs were criminogenic for London offenders who used both crack and heroin, but crime-preventive with offenders who did not use that combination of drugs. Initial findings that RJCs were criminogenic for Aboriginal Australians did not persist into the long-term follow-up.

Conducting RJ conferences Selection of police facilitators of RJCs based on innate ability is more important than experience or practice in generating procedural justice as perceived by offenders attending RJCs, which are more likely to be completed if a deadline is imposed.

Testing theoretical mechanisms (Canberra only) **Procedural Justice:** Offenders often show higher levels of perceived procedural justice after RJCs, although that does not always lead to reduced recidivism. **Reintegrative Shaming:** Offenders experienced both stigmatic and reintegrative shaming in RJCs. **Interaction Ritual:** Juveniles randomly assigned to RJCs in the Canberra property and violence experiments had less repeat offending if observers of their conferences had coded higher levels of group solidarity and reintegration.

Introduction: how does experimental evidence grow?

The modern development of restorative justice policies has arguably been an exemplar of evidence-based policymaking, both for better and for worse. Restorative justice has been better in its use of randomized controlled trials—the clearest and most valid method for testing any justice policy (Sherman et al. 1997)—from the earliest days of a global social movement to add restorative justice conferences (RJCs) to the Common Law toolkit of responses to crime. It has been worse because so much practice and governmental funding has ignored strong experimental evidence on the benefits of RJCs—especially their high value for money with serious and frequent offenders and victims of serious crime. As a result, tens of thousands of crime victims have been denied access to RJCs on the basis of evidence-free, intuitively political decisions that it does not “feel right” to use RJCs in their cases—even if it provides major reductions in post-traumatic stress symptoms of the crime victims and prevents other people from even becoming victims.

This article is primarily about the “better” side of restorative justice as an exemplar of evidence-based policymaking. Our focus is not on how knowledge gets used but on how it is generated: what we know and how we know it after two decades of testing RJCs. Our particular concern is that so many policy experiments on the same research questions get done in different ways in different

places, leaving the knowledge itself in a more uncertain state than optimal. Systematic reviews and research synthesis, while worthwhile, cannot solve the problems generated by wide differences across experiments in how they were done or what measures they used. While the evidence presented and reviewed in this article also suffers from some variations in analysis methods, all of the tests analyzed used exactly the same training and evaluation designs. We can, at least, demonstrate the feasibility of testing an identical method of dispensing justice in a uniform way across 12 experiments in two countries and four research sites.

Programs versus ad hoc experiments

The history of experimental criminology is largely a collection of stand-alone experiments. Unlike experimental psychology, in which replication attempts are frequent, if often unsuccessful (Open Science Collaboration 2015), replication attempts remain rare in experimental criminology. The dearth of replication attempts creates many problems for both theory and public policy, since un-repeated experiments have only limited scope for systematic reviews that assess the reliability and external validity of any single finding. This fact limits the potential value of research synthesis, from “What Works” reviews (Sherman et al. 1997) to the Cochrane and Campbell Collaborations (Farrington and Petrosino 2001).

The problem of *infrequent* replication of any kind is compounded by the frequency of *modified* replication attempts that vary key features of the program or outcome measurement. When key features of the interventions—or their control groups—vary between the original and subsequent versions tested, we cannot know whether different results come from different samples or different designs. Even when systematic reviews can synthesize the evidence from repeated tests, modifications in replications can challenge the idea of synthesis itself.

One solution to these problems is an alternative model of knowledge development, making greater use of coordinated research programs testing a more uniform version of each intervention. Examples of such coordinated programs in medicine include both multi-site trials conducted simultaneously (Weinberger et al. 2001) and prospective meta-analysis spread out over a longer time period (Berlin and Ghersi 2005).

This article presents a prime criminological example of a programmatic solution to the replication problem: the Jerry Lee Program of Randomized Trials of Restorative Justice Conferences. In 12 separate tests initiated between 1995 and 2001, the program delivered two sets of multi-site trials that created a prospective meta-analysis combining both sets. Working in two countries, with up to two decades of follow-up, the Jerry Lee Program tested just one version of one intervention, delivered by professionals who were trained by the same training method and trainers, associated with McDonald (2015).

The significance of the single training method was magnified in this case by the sharp contrast between the consistency of the intervention and the diversity of the responses it evoked. The intervention asked victims and offenders meeting face-to-face to discuss just three questions, but for as long as they wished. They were allowed to discuss their experiences for 10 min or 3 h, with or without tears, shouting, mumbling, anger, sympathy, boredom, or what Collins (2004) calls the structure of effective

“interaction ritual” (Rossner 2011a*, b*¹). The wide range of emotions we observed was enhanced by a sampling strategy across our 12 randomized controlled trials (RCTs) of enrolling different kinds of offenses and offenders, different kinds of victims, different degrees of social and demographic differences between offenders and their victims, different stages of the criminal justice process, and differing degrees of sanctioning severity and stigma—all subjected to the simple, consistent, single intervention. The differences in size and diversity across the nations and communities where the tests were conducted—Canberra, Australia (pop. 300,000), London, UK (pop. 8,000,000), Newcastle, Sunderland, Tyneside and other smaller northeastern English cities, and the wealthy counties of the Thames Valley (Berkshire, Buckinghamshire and Oxfordshire)—made the contrasts in individual case characteristics even more complex by adding contrasts in social contexts.

Each trained facilitator was told to ask just three questions of a well-prepared group of people, all emotionally connected to the victim, the criminal or the crime, and to insure that everyone had a chance to say all that they wanted to about each question. The questions were (1) what happened? (2) who was affected by it and how? and (3) what should the offender do to try to repair the harm caused by the crime? Like an antibiotic that is used for a very wide range of diagnoses, these core elements of restorative justice conferences were arguably delivered with a great deal of consistency across the Jerry Lee Program.

Through both systematic observations in Canberra, and narrative reports in the UK, we have good reason to believe these elements were delivered with very high integrity in the UK tests, and with less but still reasonable consistency in Canberra. While a few of the experiments were particularly challenged by low sample sizes or proportion of cases treated as randomly assigned, the insurance of 12 separate tests minimized the scientific damage from those few weak links.

A further asset of the Jerry Lee Program is the long follow-up period we have been able to achieve, possibly the longest ever for a criminological *program* of multiple randomized trials. While this asset is so far limited to the four Australian tests—which were generally not as well delivered as the UK tests—the latter are now ready for long-term follow-up by UK researchers. One aim of this article, then, is to make the case for investment in that follow-up.

The larger aim of the article is to demonstrate the potential for using programs of randomized trials to evaluate any new method for improving justice and reducing harm. The paucity of such RCT programs may be blamed on a lack of funding, a problem we must thank John Braithwaite for having solved in the early years of the Jerry Lee Program. His extraordinary vision of how to test a theory and develop a skilled practice to implement it was well-matched by his ability to build a coalition of willing funders (Strang 2012a*, b*, c*), whose diverse interests helped to insure that multiple tests would be conducted simultaneously.

Yet massive national funding may not always be necessary to create programs of multiple randomized trials, especially if a large number of communities have already decided to “try” or even “adopt” an innovation. The example of body-worn video

¹ An asterisk designates a reference that is contained in the [Appendix](#), a list of publications, not all cited here, resulting from the Jerry Lee Program of Randomized Trials in Restorative Justice. All other references are found after the Appendix in the Reference List.

cameras for police is a case in point. The initial trial led by Rialto (California) police chief Tony Farrar as a Cambridge University master's thesis (Ariel et al. 2014) quickly led to over ten completed RCTs using an identical protocol with almost identical technology (Ariel 2014). All that was needed to turn “pilots” or “innovations” into criminological experiments was a willing experimental criminologist to give free advice in exchange for massive returns of data. In a world of ideas going “viral,” the idea of randomized experimentation to test new ideas may itself be going viral. The article returns to this question in the conclusions, reflecting on how experimental criminology may be able to prosper because of the contemporary global austerity, rather than despite it.

The article begins with the “[One program, twelve tests](#)” section describing the origins and elements of restorative justice conferences (RJCs)—how they developed with RCTs, how they were produced, tracked, measured, and with what variations across 12 tests. We match that discussion with a similar description of the 12 control groups. We then describe in the “[Consent, random assignment and treatment delivered](#)” section the process of obtaining consent to random assignment, its success, and the rates of treatment as assigned. Next, we describe in “[Measuring treatments and outcomes: short and long](#)” the measurement of the treatments, and the various interviews and criminal records collected in both Australia and the UK. “[Describing treatment delivery](#)” summarizes what we have learned about what is inside the ‘black box’ of causal mechanisms by which RJCs cause victim and offender outcomes, both theoretically and empirically, using interviews and systematic observation data. “[Causal mechanisms: inside a ‘black box’](#)” begins our numbered inventory of conclusions by describing what we know about the qualitative dimensions of delivering and receiving the treatments based on observations and interviews. “[Main effect findings so far](#)” presents evidence on the “main effects” of RJCs so far on victims and offenders. “[Moderator effect findings so far](#)” presents moderator analyses of the main effects, with the “[Discussion: more work to be done](#)” section asking what we might have done or yet do, not just to increase the knowledge itself, but also to increase the extent to which knowledge gained in these experiments may be applied in practice.

One program, twelve tests

Restorative justice conferences in practice and research

The origin of the Jerry Lee Program was the fortunate coincidence of Braithwaite's (1989) theory of reintegrative shaming and the 1989 legislative reform in New Zealand that adopted restorative justice conferencing (RJC) as the core of its juvenile justice processes. These New Zealand conferences were then observed by two New South Wales Police employees: one a police sergeant (Terry O'Connell) from the small city of Wagga Wagga, the other a police trainer and former secondary school teacher (John McDonald) in Sydney. They extracted several principles for RJC from their observations in New Zealand:

- A conference is organized by a trained facilitator, who can invite anyone who is affected by a crime or its aftermath to attend
- Invited participants include victims, offenders, their friends and family

- Offenders agree in advance to “decline to deny” their commission of the crime, and to accept responsibility for causing harm, but an RJC does not depend on a formal admission of guilt
- There is no limit to how long a conference may last; 1–3 h is typical
- The conference has three phases:
 1. Offenders describe what they did; others may add details
 2. All then consider who was affected by the crime and how, including offenders; this phase is often highly emotional, sometimes with shouts and tears
 3. The final phase is a discussion and decision about what offenders can do to repair the harm the crime caused and ensure that it will not be repeated

O’Connell and MacDonald reduced these principles to the three questions posed by the facilitator in orchestrating the discussion: what happened, who was affected, and what is to be done?

By 1991, O’Connell was using this approach to divert juvenile offenders from prosecution in Wagga Wagga (after full admission by offenders of responsibility for the offense), with MacDonald promoting its use elsewhere in New South Wales (NSW). Braithwaite observed the conferences, and focused his 1992 Sellin-Glueck Award Lecture at the American Society of Criminology on how the NSW RJC implemented his theory of reintegrative shaming (Braithwaite 1989) written before RJs were adopted in New Zealand or ever used in Australia. He proceeded to recruit Sherman, Strang and others to plan a large randomized controlled trial to test the use of RJs in NSW, which were planned to be expanded across the Sydney area. In June 1993, Braithwaite, Sherman and Strang met and presented the proposal to NSW Police Commissioner Tony Lauer, who appeared receptive to the plan, at least initially.

Yet, on Christmas Eve 1993, Police Commissioner Lauer telephoned Braithwaite to say he was rejecting the plans to expand or test RJC in NSW. Strang then proposed the idea to Peter Dawson, the Chief Police Officer of the Australian Capital Territory (ACT) in Canberra, who agreed to conduct an experiment involving several types of offenses. Sherman developed the protocol for the ACT experiments while Braithwaite raised funds from multiple sources, starting with discretionary research funding of the Australian National University’s Institute for Advanced Studies. By mid-1994, a protocol was approved by the Attorney General for the ACT, Terry Connolly, with a program of training scheduled for some 500 patrol officers in how to organize and facilitate RJs. By late 1994, Strang and Braithwaite had negotiated a contract with the Australian Federal Police (AFP) that gave Australian National University (ANU) academic staff access to the criminal history information, along with approvals of the Australian Privacy Commissioner and the ANU Ethics Committee.

In April 1995, 10 weeks before the RCT was to begin, Peter Dawson was removed as Chief Police Officer of the ACT by his AFP superiors; his replacement was an acting Chief Officer who had zero or hostile interest in the project. Yet, with the support of the ACT Attorney General and the signed contract with the AFP, Braithwaite’s ANU team and Sherman proceeded to train hundreds of uniformed patrol officers to conduct RJs and to implement the experiments on schedule at midnight on July 1, 1995.

The four Canberra experiments were collectively named the “RISE project”, Sherman’s acronym for Reintegrative Shaming Experiments, in reference to

Braithwaite's (1989) theory. The offense types were selected primarily on the basis of their high volume and low-to-medium seriousness, after discussions with many officers about their willingness to refer arrestees for various crime types to be randomly assigned to avoid prosecution. The four offense types were non-domestic (and non-sexual) violent crime committed by offenders aged under 30, property crimes against personal victims and shoplifting in large stores committed by offenders aged under 18, and driving with legally excessive levels of alcohol in the bloodstream—the latter always detected by police through proactive roadblocks and random breath testing with a breathalyser—with adult offenders. The use of RJC for these offenses was unprecedented in Canberra, as well as in most of Australia.

Five years later (and several years behind schedule), the ANU posted the first report on RISE outcomes on the Australian Institute of Criminology website. These preliminary findings included a large reduction in recidivism by violent crime offenders assigned to RJC, relative to those prosecuted as usual. UK government officials soon read this report (Sherman et al. 2000*) during negotiations with the UK Treasury over a Home Office request for extra funding to develop restorative justice. Treasury had long encouraged greater use of randomized trials, so it agreed to provide £5 million for restorative justice on the condition that it be used for RCTs.

A Home Office Request for Proposals attracted several proposed quasi-experiments, but no RCTs, from UK institutions. The only proposal for RCTs came from Sherman and Strang through the Jerry Lee Center of Criminology at the University of Pennsylvania (Penn) with ANU as a subcontractor. The bid offered to build on the RISE experience in testing RJC on UK cases, using RCT designs. While several quasi-experiments were also funded, the Jerry Lee bid won the majority of the funding available.

The Penn proposal was filed on behalf of The Justice Research Consortium, a network of three police agencies (Metropolitan Police, Thames Valley Police, and Northumbria Police), in partnership with Her Majesty's Prison Service, the new National Probation Service, and the ANU's new Centre for Restorative Justice. The proposal called for a large number of RCTs on the RISE model of diversion from prosecution. That plan was quickly discarded when the Home Office said RJ could only be used as a supplement to existing conventional justice (CJ), and not as a substitute (as in RISE). The grant also required that formal consent be obtained from both offenders and victims before an RJC could be considered. These two requirements meant that all random assignment for adults required cooperation from either courts, prisons or both; they could not be conducted solely on the basis of police discretion as in Australia.

The Penn-led team therefore developed the UK RJC program in close collaboration with courts, yielding both success and failure. The success was with the (higher-level) Crown Courts, which became very cooperative and supportive of the experiments. The relative failure was with the higher-volume (lower level) Magistrates' courts, where most of the experiments had been originally planned. Both were asked to refer cases for RJC after guilty pleas had been accepted, but sentencing had not yet been pronounced. Crown Court Judges, with support from Lord Chief Justice Harry Woolf and frequent contact with London managers Sarah Bennett and Nova Inkpen, were generally willing to adjourn sentencing for 21 days in order to allow for an RJC to take place. Magistrates' court clerks were not so cooperative. While two small RCTs in Northumbrian Magistrates' Courts were eventually completed, their samples were only achieved by dogged persistence of the Northumbria Manager, Dorothy Newbury-Birch.

Our 2001 Crown Court negotiations in London proved critical to recruiting adequate sample sizes, as confirmed by the fate of a statutory authorization of pre-sentence RJC's a decade later. When the Home Office provided funding for such conferences in 2014–15, a group of Crown Court judges decided that victims would have to consent to RJC's even before a guilty plea was offered. Since RJC staff were usually not able to cite a guilty plea, or even locate the victim in time, this requirement made it almost impossible to deliver RJC's to victims of serious crimes. Hence, RJC was seen to “fail” because the Judges set it up to fail, perhaps unknowingly, but without any reference to the previously successful practice of seeking offender and victim consent only after a guilty plea has been offered and the case adjourned for a potential RJC (Strang 2015).

What emerged from all the experimental struggles in the UK from 2001 to 2004 were eight separate RCTs that supplemented rather than replaced the CJ in each category at that time. For all control groups across all 12 experiments, the treatment conditions were the standard CJ at the time for that type of offender and offense. Table 1 summarizes both the four RISE RCTs and the eight UK RCTs with their control groups.

Sample pipelines: “suction,” not trickle-flow

Each of these 12 experiments drew cases from what is technically called a sequential “trickle-flow” rather than by “single-batch” random assignment (Sherman and Strang 2010). Yet the use of the word “flow” is problematic, at least to a hydraulic engineer. The idea that cases in randomized trials emerge from a “pipeline” of referrals (Boruch 1997) implies that there is hydraulic pressure at the back end of the pipe, pushing the contents (criminal cases rather than liquids) out of the front end like a water tap. Our experience was that very little hydraulic pressure could be generated from the back end of our pipeline. What worked for us was suction from the front end, pulling whatever contents were accessible at the back end into fast forward, sometimes against the active resistance of forces blocking the pipeline.

In Canberra, the only way we received cases was from officers making arrests, 24 h per day. The protocol was for the officers to call our research officer on duty on a dedicated mobile phone number to determine whether the case was eligible. The researcher asked a standard set of questions, and recorded the case details if it was eligible. Then, the researcher opened the next numbered envelope in the random assignment sequence for the appropriate experiment and informed the officer of what the treatment should be (prosecution or RJC).

The system worked fine when officers called us to enrol cases, but they called far less often than they could have done. Despite our training some 500 officers, our project quickly went out of sight and out of mind. The RISE project was most visible in the first 2 years whenever police mounted roadblocks for random breath testing, each of which was guaranteed to catch a few offenders. When those offenders were booked, the arresting officers usually called the RISE number for random assignment of the disposition. Since the same officers made other kinds of arrests, they could easily remember the RISE project for violence and property cases as well. After the target of 900 drink-driving cases had been met, however, that experiment stopped taking new cases, so the roadblocks disappeared as a reminder for other cases.

Table 1 The Twelve RCTs and Control Groups

Jurisdiction	Setting	Offense type	Offender age	N personal victims	RJC as diversion or supplement?	Control group	No. offenders
Canberra, Australia	Police	Non-domestic violence	11 to 29	89	Diversion	Prosecution	121
Canberra	Police	Property crime	Under 18	171	Diversion	Prosecution	248
Canberra	Police	Shoplifting vs. corps.	Under 18	No (store detectives)	Diversion	Prosecution	142
Canberra	Police	Drinking and driving	Over 17	No (video of survivor)	Diversion	Prosecution	897
London, England	Police and Crown Courts	Burglary	Over 17	270	Supplement	Conventional Justice (CJ) With no RJ	216
London	Police and Crown Courts	Robbery and street crime	Over 17	137	Supplement	CJ with no RJ	125
Northumbria, England	Police and Magistrates	Nondomestic assault	Over 17	55	Supplement	CJ with no RJ	46
Northumbria	Police and Magistrates	Property crime	Over 17	55	Supplement	CJ with no RJ	71
Northumbria	Police and Youth Offending Teams (YOTs)	Assault (reprimands and final warnings)	Under 18	77	Supplement to diversion	CJ with no RJ	113
Northumbria	Police and YOTs	Property crime and other non-violent crime (Reprimands and final warnings)	Under 18	113	Supplement to diversion	CJ with no RJ	140
Thames Valley, England	Probation	Non-domestic violence	Over 17	100	Supplement	CJ with no RJ	100
Thames Valley	Prisons	Non-domestic violence	Over 17	112	Supplement	CJ with no RJ	112
Total	12			2331			1179

We made repeated attempts to motivate the officers to call us, but we were blocked by the upper ranks of the AFP. Our attempts to use the same techniques we had used in previous police experiments were repeatedly rebuffed by the hostile upper ranks. We only asked for time to communicate with the referring officers about the progress of the experiments, just as we had in Minneapolis and Milwaukee (Sherman et al. 1992) in monthly meetings, usually accompanied by beer and pretzels. But in Canberra, even a suggestion that we bring a coffee cake to a police station for an informal discussion was rejected as a “corrupt” attempt to “bribe” the police officers to alter their judgment about whether a case was appropriate in their view for the equipoise between the two conditions needed to justify the referral. Thus, for 5 years after the initial training, the AFP never allowed us to speak to groups of officers on police premises again about the value or learning from the project.

What ultimately succeeded as the “suction” of cases from the pipeline was multiple conversations one-on-one, both day and night, between ANU research staff and AFP officers. These conversations were telegraphed by relay messages through the social networks of Canberra police, years before Facebook or other electronic social media had even been contemplated. By Strang and her team cultivating, one-on-one, a small band of supporters within the police force, the research team kept the cases coming in until all four experiments had at least 100 cases.

In London, the challenge of creating suction was even more daunting. After unsuccessful efforts to gain case referrals from both defense attorneys and court clerks, Inkpen and Bennett developed a relationship with probation officers who tracked requests for pre-sentencing reports. Had the experiment been done a few years earlier, there would have been a far higher volume of such requests, especially from Magistrates’ courts. But by 2002, cost-cutting had greatly restricted the number of pre-sentencing reports that could be done, limited to the most serious offenses, which were usually sentenced in Crown Courts. Thus, the London experiments gained the names of offenders for whom the clerks had requested pre-sentence reports, usually within 24 h of the request.

Bennett and Inkpen developed such close connections with London Probation that they were approved for training and official access to the case management systems, totally relieving the probation staff from any work on the project. Each day the Jerry Lee Program’s London team checked the details of each new guilty plea for eligibility, forwarding the eligible cases to the police officers in the RJ Units.

The RJ Units immediately assigned a police constable to contact the offender, usually by going to the prison where they were being held on remand, in order to seek the offender’s consent to meet with the victim. Once the offender agreed, the same constable approached the victim to propose a 50 % chance of meeting with the offender. If the victim agreed, then the constable used a special local number to telephone a University of Pennsylvania research officer in Philadelphia who would re-screen the case for eligibility and issue the random assignment when appropriate. By 2002, Barnes had converted the process of random assignment (for all 8 UK experiments) to a secure computer program, with an algorithm generating an instant determination of whether the victim would be offered an RJC. The constable (or other facilitator requesting the victim’s consent) immediately informed the victim of the assignment, and when this was for RJC, proceeded to schedule a convenient time for the victim to come to the prison (or other location) for the meeting.

Similar processes to enrol cases were used in the other two UK sites. In Northumbria, Newbury-Birch and eight police constables worked in a fashion similar to the two London teams of similar size (one each for south or north of the Thames). The Northumbria team extracted names with eligible cases from both Youth Offending Teams (YOTs) and Probation Offices by daily faxes of names. Fax machines were set up by the Jerry Lee Program in the probation offices so that their staff could routinely fax the daily lists for pre-sentence reports to the researchers. When on some days the fax did not arrive, Newbury-Birch would call the offices before noon to press for speedy delivery.

In Thames Valley, the process was much slower, especially with prisons, the number of which expanded from 1 to 12 prisons as the experiment progressed with a grindingly slow rate of case flow. The question for the prisons was the names of people incarcerated for eligible offense types who ideally were approaching a scheduled time to be released. The probation team then attempted to obtain consent from the prisoners, and if successful, then contacted their victims for consent and random assignment (see Table 2).

Exactly what proportion of potentially eligible cases we were able to capture is difficult to determine. While Table 2 shows the cases that we reviewed in England, we could only review a sample of potential cases in RISE. Strang (2002*: 69) reports that, out of a 6-month universe of eligible arrestees for the property experiment, 12 % were referred into RISE. For the violence experiment, the rate was 11 %. What biases caused the police to refer some cases to RISE and not others remains unknown, thus reducing the external validity of the findings even within Canberra. But since the project could only proceed on the basis that officers could refer cases to either prosecution or RJC without random assignment—if they felt personally certain that the referral was exactly what was best for that case and could not ‘risk’ the case being assigned to the alternative treatment—there was little scope for capturing a larger share of the pipeline. In principle, the sample was described as cases in which the arresting officers were equally inclined to think that *either* prosecution *or* RJC would be appropriate dispositions for each arrest referred to random assignment—something for which a truly eligible pipeline could not be identified in retrospect from records alone.

In all these efforts, our team continuously promoted the “coalition of the willing” (Strang 2012a*, b*, c*) to extract by “suction” the number of cases needed for adequate statistical power in each test. What we did not do was to test or even document our success in getting conferences to happen—the number of visits to victims and offenders, phone calls to their supporters, taxi fares paid or police cars sent to get participants to RJC on time, even child care of crying children outside the meeting room. That oversight was arguably an important failure on our part, since we failed to describe the full conditions necessary to operate a successful RJC production line. Efforts to operate RJ programs in the years since our experiments have been more hampered by their challenges in obtaining cases than by any other challenge, perhaps because they were not set up on the principle of “doing what it takes” to make an RJC happen.

Consent, random assignment and treatment delivered

A major challenge to RJC is the skeptic’s presumption that victims and offenders will refuse to meet with each other, even when invited to do so by police or probation

Table 2 Pipelines and consent rates for 12 experiments

Experiment	No. cases examined	Practical and eligible	Offender contacted	Offender consented	Victim contacted	Victim consent and random assignment	Rate of R.A. per 100 offender contacts (%)
RISE violence	100	100	121	121	89	100 cases, 121 offenders	90 % ^a
RISE property	173	173	248	248	171	173 cases, 248 offenders	90 % ^a
RISE shoplift	113	113	142	142	Non-victim	113 ^b	90 % ^a
RISE drink-drive	900	900	897	897	Non-victim	987 ^b	90 % ^a
London burglary	577	422	393	301	284	169	43.0
London robbery	401	277	264	215	193	103	39.0
Northumbria adult assault	150	108	100	82	77	39	39.0
Northumbria adult property	311	213	187	106	102	54	28.9
Northumbria youth assault	220	178	168	97	95	65	38.7
Northumbria youth property and other crime	283	181	173	117	112	87	50.3
Thames Valley community sentence for violence	669	205	198	130	120	53	26.8
Thames Valley prison sentence for violence	738	501	501	313	247	94	18.8
Totals	4635	3371	3392	2769	1490	NA	NA

^a RISE offender consent based on contemporaneous estimates; no data recorded

^b Victim consent not applicable in nonvictim RISE experiments

officers. While consent to RJC is hardly universal, our evidence shows it was far higher than sceptics presume. Yet, it also seems that more formal processes of seeking consent (as in the English experiments) yield lower take-up rates than less formal processes (as in RISE). Had there been a requirement for formal consent by both parties in each case, the Australian experiments may never have been completed.

Yet, for all the ease of getting the cases to random assignment in RISE, the police capacity to get the RJC to occur was much better in England. This section describes the success of RISE in consent against less success in treatment-as-randomly-assigned, with less success at consent in England but far higher rates of treatment as assigned.

Take-up rates by victims and offenders

RISE The RISE project handled consent informally. Arrestees in eligible cases with victims were simply asked by police, while they were being booked, whether they would be happy to have a meeting with the victim rather than being prosecuted in court. Almost 100 % said yes, enabling the arresting officer to call the RISE staff any hour of the day or day of the week for random assignment of treatment. This offer to offenders was made even more attractive in RISE because it meant the offender could avoid a criminal record. Victim consent was obtained in RISE only after the case was randomly assigned to a designated RJ officer who would organize the RJC; it was that RJ officer who would call the victim to ask them when (not whether) they would like to meet with their offenders. On that basis, Strang estimates that some 90 % of the personal victims invited to attend a conference agreed to do so. The larger problem in Canberra was that such a high proportion of cases assigned to conference never received a conference. In the violence and property experiments combined, 23 % of the personal victims assigned to an RJC never actually attended one because it never took place (Strang 2002: 81).

London In London, the Jerry Lee Program tested RJC's with some of the most serious cases of the 12 experiments, in which both offenders and victims had some reluctance to meet. Some of the victims had been seriously injured by their offenders in stranger robberies; one taxi driver was hospitalized for over a week. The robber was initially reluctant to accept a 50 % chance to meet with his victim, although he agreed to do so—as four-fifths of the offenders did when asked by police (Table 2). When randomly assigned to attend a conference, the robber spent most of the RJC weeping apologetically and saying he had not meant to hurt the victim so badly. Similarly, almost half the burglary victims had seen the offender in their homes. Nonetheless, over half of all burglary victims agreed to random assignment for a meeting if their offender had consented first (Table 2).

Other UK sites The UK experiments described in Table 2 suggest a pattern of lower consent rates for adult post-sentencing cases than for other adult crimes—but this may be due to institutional differences and to the seriousness of the crimes, rather than stage of the criminal process. Three of the four joint offender-and-victim consent rates for pre-sentence adult cases in London and Northumbria were about 40 %, with only the Northumbria property crime experiment as low as 30 %. But the two post-sentence violence cases in Thames Valley, with perhaps more serious victim injuries, had an average of 21 % joint consent, about half as high as the other adult RCTs. (The two

youth experiments in Northumbria are not comparable, since they were merely comparing two different ways of diverting young people from prosecution on first and second offenses only; parental consent was an additional requirement not found in adult cases.) To our knowledge, this is the only systematic evidence that take-up rates are higher for pre-sentence than for post-sentence offers of RJC.

Treatment delivered as assigned

All 12 RCTs faced challenges in implementing the treatment as assigned (TAA). If the treatment is defined as a *policy* of trying to treat people with either conventional or restorative justice, the rates of successful delivery of each policy were high. That is, people were prosecuted when random assignment said to prosecute, even though they may never have appeared in court for a wide variety of reasons, many of them administrative; but the policy for what to do with them was never altered. If the treatment is defined as implementing the *theory* of either conventional or restorative justice, then the TAA rates are much lower (Sherman and Strang 2004b*). For theorists of restorative justice, these experiments are unsatisfactory, since they include so many cases assigned to RJC that never received them. In no case, however, did that percentage drop below a ratio of at least 10 to 1 compared to the control group. Thus, even in theory, the experiments all compared groups that had very large differences in the rates at which they experienced RJC.

Using the theory-testing definition as the most conservative approach to measuring TAA, Table 3 presents the rates at which each of the experiments recorded various dispositions of the cases.

Table 3 shows that the TAA rates were substantially higher in the UK experiments (mean = 94 %) than in the Australian RCTs (mean = 86 %). For the delivery of RJC as assigned, the difference was similar: a mean of 81.3 % in RISE and 88.3 % in the UK. This contrast is largely explained by differences in organizational infrastructure between the policing arrangements for RJ in the two countries. In the RISE tests, both infrastructure and leadership suffered recurrent changes. At various times, a special “diversionary conferencing” unit was created, changed and re-created to manage the process of delivering the RJC, both within and outside of the random assignment sample. No one leader was held accountable for the cases, let alone the results. Facilitators for the RISE conferences at some points were full-time specialists; at other points, they were general patrol officers who had attended the training but had no prior experience in facilitating an RJC. The average number of previous RJC for the facilitators in the three juvenile experiments was under five for the first 3 years of RISE, but went up substantially in the last 2 years when the cases were concentrated in a specialist unit.

The UK experiments, in contrast to those in RISE, were led by the same strong operational staff with a single organizational structure from start to finish. The six UK police experiments all operated with a full-time specialist model, vertically integrating the tasks in each case from offender consent to facilitating the conference and following up with victims on promises made. The prison and probation experiments used a more flexible staffing model, but almost all of them stayed with the project for 4 years and acquired substantial experience. This level of stability helped to avoid the kinds of

Table 3 Delivery of RJC and CJ as randomly assigned in 10 police-led experiments^a

Experiment	No. cases assigned to RJC	No. cases treated with RJC	No. cases assigned to CJ	No. cases treated with CJ	Percent of RJC cases treated with RJC	Percent of CJ cases treated with CJ	Full sample treatment as assigned
RISE violence	50	41	50	45	82.0 ^b	90.0	86.0
RISE property	86	56	87	77	65.1 ^b	88.5	76.9
RISE shoplift	59	52	54	48	88.1 ^b	88.9	88.5
RISE drink-drive	450	405	450	432	90.0	96.0	93.0
London burglary	118	96	109	109	81.4	100.0	90.3
London robbery	61	51	67	67	83.6	100.0	92.2
Northumbria adult assault	24	21	21	21	87.5	100.0	93.3
Northumbria adult property	34	31	36	36	91.2	100.0	95.7
Northumbria youth assault	51	46	50	50	90.2	100.0	95.1
Northumbria youth property	54	52	56	56	96.3	100.0	98.2

^a Data not presently accessible for the two probation-led experiments in Thames Valley

^b Data differ from Sherman and Strang 2004b fig. 2, based on persons in cases randomized; this table uses case-based random assignment for RISE (identical for drink-drive experiment only) and person-based analysis for the UK experiments

problems that emerged in Canberra, where several cases were assigned to constables who never even tried to arrange an RJC. When offenders failed to appear for RJC in the UK, the police would locate them and re-schedule the conference—another difference from Canberra, where RJC were often dropped after an offender failure to appear, and the case referred to prosecution.

Even our role as criminologists was different in the two countries. In RISE, we were the arm's length evaluators, with no role in generating cases or implementing random assignment. In the UK, we were tasked by the funders and the police with insuring the best implementation of the project so that others could evaluate it. That meant our full-time site managers were the primary people responsible for obtaining consent and delivering RJC, in equal partnership with the dedicated agency staff who performed the front-line work. Whether this was all structural, however, depends on whether we had learned enough from watching the AFP in RISE to do a better job at case management in the UK. What we learned should probably be spelled out in an operational manual, as we discuss below in “**Discussion: more work to be done**”.

From a theory-testing standpoint, the most problematic of the 12 experiments is the juvenile property crime RCT in Canberra, where the percentage of cases in which RJC actually occurred after random assignment was only 65 %. Nonetheless, the percent of cases assigned to prosecution in which RJC were delivered was only 1 %. Thus, the intent-to-treat (ITT) analysis of these cases as randomly assigned maintains strong causal inference of about different outcomes from very different rates of RJC delivery, or 65 times more RJC delivery in the ITT group for RJC than for prosecution.

The estimates for the benefits of RJC's resulting from these 12 experiments may substantially under-estimate what could be obtained in theory. Yet, they are arguably more useful as estimates of the effectiveness of a policy in practice, as opposed to its underlying "efficacy" under conditions of perfect compliance.

Measuring treatments and outcomes: short and long

The Jerry Lee Program of Randomized Trials in RJC's has a rich, if not entirely consistent, set of measures of both treatment delivery and outcomes. The Program remains a work in progress. Outcomes have been reported for all 12 RCTs for up to 2 years, although 6 of the UK trials have only reported outcomes for the partial sample gathered by an independent evaluator within its own reporting deadline. Treatment delivery has been fully analyzed in 8 of the program's 12 RCTs, but further analysis of the full sample has yet to be completed in 4 of the UK tests. In the 4 RISE RCTs, the detailed systematic observation of both RJC's and control cases has provided a rich theoretical analysis of RJC's for three kinds of theory: reintegrative shaming, procedural justice and interaction ritual chains. RISE also has the benefit of 10-year interviews with hundreds of offenders and victims, as well as up to 18 years of mortality data and criminal history records, post-random assignment, for both victims and offenders, for which analysis is in progress.

The eight UK experiments, in contrast, collected much less qualitative measurement of treatment delivery than RISE. Nor have the UK tests had any follow-up data collection since 2007. Yet, they offer a far wider range of samples than the RISE tests, across different offense types and different points of the criminal justice system.

These differences in measurement between the Canberra RISE and UK parts of the Jerry Lee Program were created by external constraints of the funders. The Jerry Lee Program was created by a merger of the existing RISE project with the new UK project, created when the Jerry Lee Center of Criminology at the University of Pennsylvania won the Home Office grant to conduct the eight UK experiments. The Home Office grant required the Jerry Lee team to play a different role in England from the role we had played in Australia. In Canberra, we had served as both "developer" and "evaluator" of the RJC program (Eisner 2009; Sherman and Strang 2009a*). In England, by government policy, the two roles had to be separated. While the University of Pennsylvania and the Justice Research Consortium had won the grant to develop the program, the University of Sheffield was selected as the independent evaluator of its effects. That meant that while Penn would recruit all cases and document their random assignment, all post-treatment impact analysis funded by the Home Office was assigned to Sheffield (see all reports by Shapland et al.*).

Thanks to funding from the Jerry Lee Foundation, George Pine and other philanthropists, the Jerry Lee Center of Criminology at the University of Pennsylvania was also able to fund its own data collection on UK outcomes. This had several advantages, but it led to a somewhat confusing situation with several features:

- In the UK RCTs, substantially more cases were randomly assigned than the University of Sheffield had funding to gather data on within its grant budget and time frame, leaving a larger sample size unanalyzed for the official government

reports, even though the full sample has been used in some other analyses (e.g., Bennett 2008*).

- The eight UK RCTs were merged into seven in the Shapland et al. (2008*) reports because the independent evaluator chose to combine the two juvenile RCTs in Northumbria, which pooled property crime–other and violent offenses.
- Systematic observations of court appearances and conferences, as well as victim and offender interviews, were attempted for all cases in RISE, but only for partial samples of cases (by Shapland’s team) in the eight UK experiments.
- The London victim interviews conducted by Angel (2005*; Sherman et al. 2005*; Angel et al. 2014*) were focused primarily on measuring post-traumatic stress symptoms and other specific items, and were not linked to other measures of victim outcomes collected for the Sheffield sample.
- RISE attempted to obtain detailed interview measures of offender perceptions of procedural justice and other attitudes within 6 months, 2 and 10 years of the random assignment; the UK experiments did not.
- UK RCTs as analyzed by Shapland et al. (2008*) and summarized by Sherman and Strang (2012*) provide precise cost-effectiveness estimates of the investment in the RJs; RISE did not.

These RISE versus UK differences are especially pronounced in terms of offender recidivism outcome measures. All 12 of the Jerry Lee Program’s RCTs have reported findings on offender recidivism (Shapland et al. 2008*; Sherman et al. 2000*; Sherman and Strang 2012*; Strang et al. 2013*), but in only one analysis were identical measures used for all 12 (Sherman and Strang 2012*)—and even that one relied on Shapland’s combination of results from the two Northumbria juvenile RCTs. The eight UK tests are largely limited to 2-year after-only reconviction rates from the Shapland et al. (2008*) independent evaluation of the UK RCTs, with the truncated sample of all randomized UK cases (but see Bennett 2008*). The after-only approach is arguably not as strong as the before–after, difference-in-difference approach, which has been reported for at least 2 years before and after random assignment for RISE (Sherman et al. 2000*; Woods 2009*). This approach better adjusts for the baseline differences in offending rates between experimental and control groups. Because many of the sample sizes are relatively small, the difference-in-difference approach helps to improve the precision of the estimated effects.

RISE recidivism outcomes are also reported for much longer time periods than for the UK experiments. This reflects both differences in funding and in the complexity of compiling the criminal history data from the partner police agencies, which has been far easier with the single RISE partner (the Australian Federal Police) than with the three separate UK police partners. The long-term data in Australia have been especially important in clarifying the effect of RJs on juvenile Aboriginal offenders, as reported below. Arguably the greatest gap and most pressing agenda for the UK experiments is to obtain follow-up measures of recidivism for as long as RISE has.

Finally, the RISE recidivism data have been able to distinguish different kinds of offenses in ways that have been more challenging for the UK experiments. While Shapland et al. (2008*) computed the estimated cost of the various offense types included in offender recidivism, their evaluation did not clearly distinguish between *new* offenses against victims from either *breaches* of previous sentencing orders

(technical violations) or *non-victim* offenses, such as possession of illegal drugs, commercial burglary, or drink-driving, as Woods (2009: 47*) did for RISE.

Describing treatment delivery

The qualitative dimensions of treatment delivery in the four RISE experiments were measured with a systematic observation instrument available online at the University of Michigan ICPSR (see <http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/2993?geography=Global>). The global ratings of the theoretical dimensions of the RJs and the control group court appearances were tested for inter-rater reliability early in the first year of RISE, with high reliability scores (Harris and Burton 1997*, 1998*). The data taken from these instruments have been used in analyses by Rossner (Rossner 2008a*, b*, 2011a*, b*, 2013*) as reported below. In addition, Harris (2000*, 2001*) and Braithwaite and Braithwaite (2001*) have analyzed these data, while Inkpen (1999*) has reported an ethnographic study of a sample of the same conferences.

Facilitator differences in procedural justice

Individual differences across the many officers in Canberra who facilitated RJs were identified from analysis of the systematic observation data (Sherman et al. 2003*). Woods (2009*) then used the initial post-RJC interview data with offenders, aggregated for each facilitator, to test three hypotheses to explain facilitator differences in the three RISE experiments involving juveniles (i.e., this analysis omitted the drinking-driving experiment). Woods' analysis was designed to discover whether there was any difference in interview-measured perceptions of procedural justice by offenders who sat through the conferences according upon the a) *total experience*, b) *recent practice*, or c) *innate ability* from the first RJC of each police officer facilitating the conference. In effect, he compared two "practice makes perfect" hypotheses to one "natural ability" hypothesis (Gladwell 2008). He found no evidence of any improvement in procedural justice perceptions as individual facilitators gained more practice or had recent experience. He did, however, find a large and consistent difference between some facilitators and others in offender perceptions of the officers' fairness. This leads us to the following:

Conclusion #1: Selection of facilitators based on innate ability is more important than experience or practice in generating procedural justice from restorative justice conferences.

That conclusion notwithstanding, the research so far does not tell us how to predict whether one potential facilitator has more ability to generate procedural justice than another. It just tells us that this difference can be measured in practice based on offender interviews. Even that finding may have more general applicability to the selection of police and others exercising authority in the justice systems.

Process issues in completing conferences

Woods (2009*) also examined the differences between RJ conferences in RISE that were completed and those that ultimately failed to occur, for both administrative

reasons (such as an officer deciding not to carry through or forgetting about the task assigned) and offender-related reasons (such as the offender moving out of the jurisdiction or being arrested on a new charge). Woods found that while attempts to hold RJC shortly after the arrest had a high risk of failure, so did those that were delayed beyond several months. This leads to the following:

Conclusion #2: RJ Conferences with both victims and offenders present were most likely to be completed in Canberra if they were scheduled to take place roughly between 6 and 12 weeks after random assignment.

The Canberra evidence, however, must be read in light of the diversion of the RJ case from prosecution, which meant that the timing of an RJC lacked any court or other deadline. The UK evidence for pre-sentence RJC, in contrast, was strongly tied to court deadlines. With 21 days adjournment from guilty plea to sentencing in Crown Court, for example, there was a strong sense of urgency by all parties in completing an RJC process before the sentencing hearing. Similar deadlines were present with the prison, Magistrates' Court and juvenile experiments in the UK, although not with the post-sentencing probation experiment under community sentencing or the prison experiment. The UK experiments, as Table 3 shows, had higher treatment as assigned rates than in RISE. While there were many other organizational differences between the RISE and UK experiments, there is still reason to believe this evidence supports the following assessment:

Conclusion #3: The urgency of a sentencing or other court deadline may lead to higher rates of completed conferences from using RJC as a supplement to court rather than using it as a substitute.

Causal mechanisms: inside a 'black box'

Randomized experiments are often criticized for not testing causal mechanisms that may explain any effects of different treatments. While this criticism fails to acknowledge the long history of science providing unexplained benefits based solely on effects—such as the prevention of scurvy with citrus fruit or the prevention of cholera with clean water (Sherman 2015)—there are undoubted advantages to understanding plausible causal mechanisms for clear effects. Funding differences allowed more investment in this task in RISE than in the UK, about which very little evidence is available concerning the 'black box' of causation across all cases (but see all Shapland* reports for selected samples of cases). Whether the causal effects found in RISE would be valid for the UK experiments is unknown, but it is at least possible that they are more generally present in RJC.

Offender perceptions of procedural justice

There is strong evidence from RISE that RJC increase offender perceptions of procedural justice, at least when used as a diversion from court. Barnes (1999*) first found this, in a theoretically coherent analysis of both process and outcome variables, in

the RISE drink-driving experiment (see also Barnes et al. 2015*). Tyler et al. (2007*), using a more multivariate strategy with similar data, found the same result. Strang et al. (2011*), in the RISE Final Progress Report, also reported item-by-item differences that showed higher levels of perceived procedural fairness in all four of the RISE experiments among the RJC-assigned offenders compared to those assigned to court.

Conclusion #4: Offenders in all four RISE experiments showed higher levels of perceived procedural justice if they were randomly assigned to (but not necessarily completed) RJ conferences than if they were assigned to court.

Shaming: reintegrative and stigmatic

The RISE experiments accomplished their primary theoretical purpose of testing Braithwaite's (1989) theory of reintegrative shaming, which RISE generally confirmed but elaborated. The fundamental hypothesis was that RJC's would produce higher levels of reintegrative shaming (hate the sin but love the sinner) and lower levels of stigmatic shaming (hate the sinner and the sin) than prosecution in court. As predicted, the RISE experiments all produced much higher levels of reintegrative shaming in perceptions of offenders assigned to RJC's than among those assigned to prosecution. *Not* as predicted, however, the RJC's also caused offenders to feel more "disapproved of" than similar offenders said they had felt in court (Harris 2001*: 130). These findings led to substantial revisions of reintegrative shaming theory with more complex conceptualization of shame and guilt, drawing on the nuanced measures in both the RISE observations and offender interviews (Braithwaite and Braithwaite 2001*).

This evidence lends support to the reformulation of the idea that reintegrative shaming is the opposite of disintegrative (or stigmatic) shaming (Braithwaite and Braithwaite 2001*). Braithwaite's (1989) initial theory had presented the two on a continuum from stigmatic to reintegrative. But the RISE evidence suggested that the offender's experience could be classified on two independent dimensions simultaneously, such as being high on *both* stigmatic and reintegrative shaming, so that shaming of both kinds can co-exist in ways that enhance an offender's sense of guilt. On this basis, RJC's producing a higher level of feeling disapproval may still act as a crime prevention mechanism. Thus Harris's (2001*) RISE analysis, although limited to the drink-driving experiment in which RJC's failed to prevent crime, lends support to a revised view of the theoretical mechanism of RJC's can create more fulsomely than a brief court appearance, along with higher levels of procedural justice, even though RJC's failed to reduce crime in that experiment:

Conclusion #5: RJC's for Canberra drink-driving offenders produced a higher level of shame and guilt than court appearances, even though they reported a higher perceived level of procedural justice, with no reduction in recidivism.

Further insight into the causal mechanisms of the RISE offenders' experience in RJC versus prosecution, by experiment, were gained from our 10-year follow-up survey (Table 4). The response rates were of borderline utility, especially in the shoplifting experiment, which was also the only one to have a higher response rate from the

prosecution-assigned group than from the RJC group—and the only one where less than half of the respondents could even remember the experience. We present these results for whatever readers may think they are worth. Even with possible sampling bias, it is interesting that they show such strong effect sizes on so many items. The effect sizes on offenders being pleased that their cases were handled by RJC (compared to prosecution) were very large, as were the effect sizes on making them less angry and bitter about the justice they received. Shame over the crime or themselves was not very different between treatment groups, thus undermining (but not falsifying) any kind of shaming theory of recidivism prevention. Only the violent offenders said the RJC experience was a turning point in their lives, which fits the fact that the strongest reductions in frequency of repeat offending in all 12 experiments were in the RISE violence experiment, where the emotions at play tended to be much more powerful than in the other three RISE tests.

Whatever the causal mechanism may be, the survey shows striking persistence of differences between the RJC-assigned and prosecution-assigned groups. Whatever happened in the RJCs, it appears to have been highly memorable and affecting, at least compared with any other hour or two in most people's lives.

Conclusion #6: The offenders' experience of RJC-assignment in RISE, at least among respondents to a 10-year survey, produced lasting differences in attitudes and emotions from those of prosecution-assigned offenders who responded to the survey, almost all showing better self-reported re-offending than the prosecution group respondents.

Interaction ritual theory

One theory of an RJC's causal mechanism was published after RISE began—and indeed was partly shaped by RISE itself: Randall Collins' (2004) reformulation of Erving Goffman's interaction ritual perspective. Citing RISE (among much other evidence), Collins proposed that the key elements of a successful interaction ritual are a) co-presence of all participants in the same place, excluding non-participants; b) a shared focus on a particular topic; and c) a conversational and bodily rhythm; all of which recommit all those present to the shared morality of a group. He stated this in terms of linear dimensions, a continuum by which ritual encounters can vary in the degree to which they produce the key elements of the theory. The more successful they are in doing so, Collins suggests, the greater the level of group solidarity, emotional energy, and recommitment to the shared morality.

The work of former Jerry Lee Program London staffer Rossner (2008a, b*, 2011a, b*, 2013*) has tested Collins' theory with qualitative analysis of a sample of RJCs in the London experiments and quasi-experimental analysis of the observational data of the RJCs in the RISE youth violence and youth property experiments. While she could not compare the RJC cases to the court cases, she was able to analyse the differences across conferences in three key elements of Collins theory: reintegration, solidarity and emotional energy. Moreover, Rossner (2013*: 131–135) could relate the success of each RJC in achieving a strong interaction ritual to both the prevalence and frequency of rearrest. She shows that higher levels of solidarity and reintegration in an RJC

Table 4 Ten-year survey of RISE offenders, by experiment

Experiment:	Drink-drive	Violence	Property	Shoplifting
RJC response %	61 %	64 %	62 %	52 %
Prosecution response %	62 %	48 %	58 %	67 %
% Both groups who remembered RISE case “well”	74 %	65 %	56 %	45 %

Questions	Cohen’s D	Cohen’s D	Cohen’s D	Cohen’s D
	+ for RJC	+ for RJC	+ for RJC	+ for RJC
Were they pleased with the way their case was handled?	+1.65	+1.28	+0.72	+1.33
Should the government make RJs more widely available?	+0.67	+0.30	+0.35	+0.13
Were they ashamed of the crime they had committed?	+0.12	+0.26	+0.21	+0.03
Were they ashamed of themselves for committing the crime?	+0.08	+0.30	+0.22	+0.18
Did the punishment they received make up for the harm the crime had caused?	+0.54	+0.71	+0.13	+0.7
Were they now angry about the punishment they received?	-0.50	-0.48	-0.04	-0.24
Were they now bitter about the punishment they received?	-0.43	-0.26	-0.01	-0.05
Did they currently want to get back at those who caused their punishment?	-0.29	-0.28	-0.33	+0.36
Was their experience an important event in their lives?	+0.15	+0.28	+0.21	+0.03
Was their experience one that affected their lives?	+0.12	-0.23	+0.17	+0.21
Was their experience a turning point in their lives?	+0.16	+0.50	+0.07	+0.07
Did the experience help you to obey the law?	+0.32	+0.25	+0.25	+0.34
Did the experience help your friends and family to obey the law?	+0.55	+0.02	+0.17	+0.55

predict lower levels of reoffending, but that higher levels of emotional energy do not. Her findings were somewhat complicated by other factors, such as the presence or absence of a prior record. Nonetheless, her evidence provides correlational, if not causal, support for this assessment:

Conclusion #7: Juveniles randomly assigned to RJs in the RISE property and violence experiments had less repeat offending when observations of their conferences showed higher rather than lower levels of solidarity and reintegration.

Main effect findings so far

Offender recidivism

While RJs in general proved more effective than CJ in preventing recidivism, the Jerry Lee Program has found important complexities in both short-term and long-term results. In two of the four RISE experiments, for example, the after-only rate of convictions was higher for the offenders randomly assigned to receive RJs than it was for the cases assigned to prosecution. Both the drink-driving and the juvenile

property crime experiments appeared to backfire by this measure, causing more crime rather than less in the first 2 years of follow-up. Other complexities of RJC effects on recidivism are related to a) whether there is a personal victim who can be included in the RJC, b) the use of cost (or “harm”) of crime rather than counts of crime as if all crimes are created equal, and c) the length of follow-up period in which effectiveness is defined.

Personal victim offenses There are strong theoretical reasons to believe that because an RJC without a personal victim cannot be restorative to a victim, it cannot really be an RJC. Whether this is true as a normative theory—and there are plausible claims that it is possible to ‘construct’ victims, including members of the offender’s family, for the purpose of an RJC—an empirical model of causation that draws on empathy for another human being as a victim cannot be achieved in RJC without a victim (Sherman and Strang 2011*). Offenders, in theory, cannot experience as much intensity of remorse without someone they had actually harmed expressing the pain they had suffered. It is for that reason that the RISE and UK experiments with personal victims present in the RJC (RISE without the shoplifting and drink-driving experiments) were the only Jerry Lee Program experiments incorporated into our Campbell Collaboration Systematic Review of RJC effects (Sherman 2014*; Strang et al. 2013*). That review found overall reductions in recidivism of RJC compared to conventional justice, for RJC with personal victims present. (The exclusion from that review of two RCTs on offenses without personal victims was also made on policy grounds, since victim benefits are an equally if not more important aim of RJC, and they are impossible to achieve with RJC for non-victim crimes.)

Conclusion #8: The average effect of RJC on offenders is to reduce the frequency of repeat offending, as observed in 9 out of 10 experiments with personal victims. One of two experiments without a personal victim (drink-driving) showed an increase in frequency of repeat offending.

Cost of repeat offending Most experimental criminology counts repeat offending as if all crime is created equal. It is not (Sherman 2007, 2013). The use of a crime harm index (CHI) that weights each crime with a ratio-level indicator of seriousness is a far superior approach to examining the effects of any justice policy. The independent evaluator of our UK experiments (Shapland et al. 2008*: 64) used just such an approach in testing the cost-effectiveness of RJC in our UK RCTs. Their method used the Home Office data on the costs of crime (to both victims and government) to compare the financial value of crimes prevented by adding RJC to Conventional Justice vs. the costs of providing RJC in our UK experiments.

This calculation was extremely important, but not readily transparent to many readers. We therefore re-computed the cost–benefit ratios from data in the Sheffield report, spelling out the overall cost–benefit ratio for the UK experiments at 8:1, or £8 in costs of crime prevented for every £1 spent on providing RJC to supplement the prosecution and sentencing (Sherman and Strang 2012*; Strang et al. 2013*). This ratio ranged from a high of 14:1 in the London robbery and burglary cases combined, to a low of 1.2:1 in all Northumbrian cases combined, with a majority of them juvenile

offenses. The benefits were highest where the frequency and seriousness of prior offending was highest, in the London experiments, where the burglary offenders had a 5-year pre-conviction mean of 5.89 prior burglary convictions and the robbery offenders had a mean of 3.48 prior robbery convictions (Bennett 2008).

Conclusion #9: RJs were cost-effective in all seven UK tests preventing more cost of crime in the short run of 2 years follow-up than the cost of delivering the RJs, with far more cost-effectiveness among serious offenders with many prior convictions.

Similar cost-effectiveness estimates are not available for the RISE cases.

Short- or long-term recidivism effects The Jerry Lee Program has highlighted a central issue in evidence-based policy: how long is long enough, or too long, to measure outcome differences between treatments? While various authorities have recommended a 2-year *minimum* follow-up of any program randomly assigned to individuals, there is currently no discussion of a *maximum* period for follow-up. While our analyses show clear overall effects of RJs on reducing recidivism at 2 years (Strang et al. 2013), the RISE analyses show that these benefits have disappeared by 15 years (Sherman et al. 2015a, b*). These data suggest the following assessment:

Conclusion #10. While RJs reduce recidivism for 2 years, analyses of the RISE evidence to date shows no main effect on recidivism after 15 or more years.

Victim benefits

Short-term victim benefits The impact of RJs on victims has been highly beneficial in both RISE and the UK experiments. Some of these findings have been quasi-experimental, before–after differences with the group of victims who attended conferences (Strang and Sherman 2003*; Strang et al. 2006*). The most important differences, however, have been based on experimental estimates (Angel 2005*; Angel et al. 2014*; Strang 2002*; Sherman et al. 2005*).

It was Strang (2002*: 97) who first showed that RJs reduced the percentage of victims of violence and property crime who feared that the offender would revictimize them, from 18 to 5 %. More importantly, she showed that RJs reduced victims' desire for violent revenge (Strang 2002*: 138–139) against the offenders, from 20 to 7 % (and from 45 to 9 % for victims of violent crimes only) (see also Sherman et al. 2005*). Finally, she found that victims were more likely to be pleased with the way their case was dealt with if their offenders had been assigned to RJs (69 %) than if they had been prosecuted (48 %).

Conclusion #11: Victims assigned to RJs in RISE were less fearful of repeat attack by the same offenders, more pleased with the way their case was handled, and less desirous of violent revenge against their offenders than controls.

UK tests Short-term victim benefits of RJC were somewhat weaker in the UK evidence than they were in the RISE experiments. Shapland et al. (2007*: 42) found slightly weaker effects in the UK experiments, when RJC only supplemented the CJ process, rather than substituting: 72 % of RJC-assigned victims were satisfied or very satisfied compared to 60 % of victims whose cases did not receive RJC. But the UK control group (CJ) victims (unlike the RISE CJ victims) had all expressed a willingness to meet with their offenders prior to random assignment, and had often reported disappointment to the constable who obtained their consent about their not being selected for RJC.

The Campbell Systematic Review (Strang et al. 2013*) also incorporated the findings of eight sets of victim interviews by Strang and Angel, as first reported in Sherman et al. (2005*); victims were far more likely to receive apologies in RJC than in conventional justice; the RJC-assigned victims were more likely to receive apologies they found to be sincere; they were no less likely to blame themselves for the crime than conventional justice-assigned victims; in the London experiments, the RJC-assigned victims were more likely to forgive their offenders than were the CJ-assigned; and across all eight results, victims were less likely to want violent revenge if they had been assigned to meet with their offenders than if not.

Conclusion #12: Victims assigned to RJC in both the UK and RISE were more likely than control group victims to receive offender apologies, be more satisfied with their justice, and less desirous of violent revenge than controls.

The most powerful evidence of victim benefit from RJC is the Angel et al. (2014*) evidence that RJC reduce the post-traumatic stress symptoms (PTSS) reported by victims. Using a standard psychiatric diagnostic tool in telephone interviews of 192 London victims of robbery and burglary, the Angel team found 49 % fewer victims suffering clinical levels of PTSS among the RJC-assigned victims than among the victims assigned to CJ only. These findings were limited to short-term impact, but they reflect basic life functions such as sleep and ability to leave the home to go to work. They also imply a possible long-term reduction in an otherwise elevated risk of premature mortality, which has been associated with chronic PTSS, even at low levels (Kubzansky et al. 2007).

Conclusion #13: London robbery and burglary victims assigned to RJC suffered much less post-traumatic stress than controls.

Long-term victim benefits

The evidence so far shows that victim benefits of RJC last longer than any effects on offender recidivism. While our only long-term victim effects data so far come from a 10-year post-random assignment survey for the RISE violence and property experiments, Strang's (2011*) research team on this survey achieved a substantial panel response rate of 81 % ($n = 188$ out of 232 initially interviewed), which was 72 % of 260 initially sought for interviews. After 10 years, the benefits for RJC-assigned victims

remained clear: they still had half as much *anxiety* about being revictimized as victims whose cases had been prosecuted (22 % RJ vs. 44 % court, $p = .00$); half as much *anger* about the crime (58 % RJ vs. 26 % court disagreed that they were still angry, $p = .01$); and half as much *feeling of bitterness* about offense (75 % RJC vs. 38 % court disagreed that they still felt bitter, $p = .00$).

Other benefits for RJC-assigned victims, if borderline in statistical significance, were less general *fear of crime* (22 % RJC vs. 34 % prosecution, $p = .11$), and more disagreement that they would do some *harm to offender now* (80 % RJC vs. 63 % prosecution strongly disagree, $p = .10$).

Two measures that showed no difference between RJCs and court were (1) whether the treatment of their case had *put their minds at rest* (around 75 % of both RJC-assigned and prosecution-assigned said it had not) and (2) whether the victims felt *forgiveness* of the offender (20 % of both treatment groups remained unforgiving). But another, more subtle measure showed an important benefit for the RJC victims, who were more likely to have forgotten just what happened in the justice process they attended (47 %) than court-assigned victims who attended court (33 %).

Conclusion #14: Substantial victim benefits in reducing the emotional impact of the crime resulted from random assignment to RJCs in the two Canberra RISE tests and persisted for at least 10 years after the arrest of their offenders.

Moderator effect findings so far

One strength of the Jerry Lee Program has been its capacity to detect important moderator effects: not just whether RJCs “work,” but *for whom* they work more or less well, or even make things worse. Such differences have been found to date for victim gender, offense severity, offender baseline offending frequency, offender drug use, and initially for race in Australia (Strang and Sherman 2015*), although the latter appears to have disappeared in a 15-year follow-up (Sherman et al. 2015a, b*) and will be reported in detail in a separate article.

Post-traumatic stress reduction and gender

If restorative justice were to be rationed on the basis of the greatest benefits it produces for victims, there is good evidence for prioritizing women. The Angel et al. (2014*) analysis of the post-traumatic stress symptoms reduction in London showed that while RJCs reduced PTSS as a main effect, women victims had much higher PTSS levels after burglary and robbery victimizations than male victims did. They also showed much more PTSS reduction after RJCs than men: 46 % were above subclinical levels of PTSS in the female RJC-assigned group compared to 78 % for female controls, while men only had a difference of 37 % RJC-assigned versus 45 % for controls.

Conclusion #15: Female victims of robbery and burglary in London had much greater short-term reductions in PTSS levels than male victims, although both genders showed benefits of RJC on PTSS.

Repeat offending and offense severity

The Strang et al. (2013*) systematic review of RJC effects on recidivism included a moderator analysis by offense severity. The biggest effect of any moderator in that analysis (including offender age, time at risk, use of conviction outcomes only or including arrests) was the interaction of RJC with offense severity. The concept of severity was crudely indicated by the instant case being for either a violent crime or a property crime. While one of the RCTs included in the Systematic Review was not part of the Jerry Lee Program, that RCT (McGarrell and Hipple 2007; Jeong et al. 2012) used the same trainers as all the Jerry Lee Program experiments. The overall standardized mean difference in 2-year frequency repeat offending was $D = -.163$ ($P = .001$), yet the same measure for only the three property crime-only experiments was $D = .001$ ($P = .989$). The meta-analysis of the five violent crime RCTs, however, yielded a standardized mean difference in favor of the RJC of $D = -.198$ ($P = .045$). Thus, it seems fair to say that in general:

Conclusion #16: The average effect of RJC (compared to CJ) on repeat offending across all three reported property crime experiments was nil, while the average effect of RJC across five experiments with violent crime was a modest but statistically significant reduction in the frequency of repeat offending.

Repeat offending and offender baseline frequency

Another issue in using RJC is whether it is best used only for first offenders (as often claimed), and inappropriate with high-frequency offenders since for them it is “too late”: they have become “hardened criminals.” The evidence from the Jerry Lee Program in two hemispheres shows exactly the opposite.

Both the Canberra (Woods 2009*) and London experiments (Bennett 2008*) provide consistent evidence on how RJC effects vary by baseline offending frequency. Analyses in both cities use arrest frequency over a 5-year period prior to random assignment as the baseline rate of offending. The repeat offending measure in Canberra was arrest frequency in a 5-year follow-up; in London, it was time-to-failure from random assignment (or prison release) to date of first offense resulting in arrest in the time period 2002 through 2005. In both cities, the evidence shows that RJC effectiveness appears to be curvilinear: they work best for offenders with the highest and lowest frequency of prior offending. RJC work least well for offenders with a moderate frequency of prior arrests.

Sarah Bennett’s (2008*) analysis of offender time-to-failure in the two London experiments found no statistically significant differences between the RJC-assigned offenders and those equally willing to meet with consenting victims randomly assigned to the control group. “Failure time” in Bennett’s analyses was the number of days between release from prison (or random assignment date for those not in custody) and the date of the first offense that led to an arrest (Bennett 2008*: 79). This “crime-free” period was actually longer for RJC cases (compared to controls) in both experiments (Bennett 2008*: 82), especially in the robbery experiment (522 days for RJC vs. 371 days for controls), but the differences had very wide confidence intervals (range

of error). Yet, since only 61 % of the sample offenders had any rearrest during the follow-up period ending December 31, 2005, there was substantial variation to explain.

When Bennett specified more homogeneous subgroups of the experimental samples, more than a “chance” number of subgroups showed statistically significant differences between the RJC and control groups in time-to-failure. This phenomenon may be an example of Weisburd et al.’s (1993) paradox, in which smaller sample sizes are more likely than larger samples to produce statistically significant differences because smaller samples may be less heterogeneous, with smaller standard deviations. The most important instance of this was the level of baseline frequency of arrest.

First, Cox regression results indicated that the frequency of arrests in the 5 years prior to random assignment had a statistically significant interaction effect with RJC and time to failure (Bennett 2008*: 159), in both the burglary experiment ($n=227$) and the robbery and burglary experiments combined ($P<.0001$). She defined high frequency offenders as those with a mean of over seven arrests per year at risk in the 5-year pre-random assignment baseline period. These high-frequency offenders had a mean of 94 days to first offense in the control condition, but 234 days (a 149 % increase) in the experimental condition (Bennett 2008*:160).

Second, Bennett (2008*: 160) found that London robbery offenders ($n=128$) showed the same pattern. Offenders with a baseline rate of over seven arrests per year for 5 years before pleading guilty to a robbery charge had over twice the mean survival time after random assignment to an RJC (316 days) than after assignment to CJ (140 days).

In the same experiments, however, Bennett (2008*:160) also found evidence that RJC worked better to delay repeat offending if they had the *lowest* baseline rates of arrest than if they had *medium* rates. She defined the lowest rates of baseline arrests as less than two arrests per year, and medium rates as between two and seven arrests per year, in the 5 years prior to date of random assignment. Robbery offenders with the lowest baseline rates had a mean survival time of 382 days in the control and 634 days in the RJC-assigned condition, or a statistically significant 66 % increase in time to first repeat offense (see Fig. 1). A significant increase in failure time for lowest baseline-rate burglary offenders was in the same direction, but much smaller: 507 days over 474 days (7 % more).

Bennett’s (2008: 160) London analysis also found evidence against using RJC for medium rate offenders (2–7 arrests per year in baseline). Medium baseline-rate offenders in burglary had only a 13 % increase in failure time after assignment to RJC. Even worse, medium-rate robbers had a statistically non-significant, but *backfiring* effect from RJC—which cut their mean time to failure from 350 days for controls to 219 days for RJC (a 37 % reduction, or a 60 % benefit from *not* using restorative justice).

Daniel Woods’ (2009*) analysis of the three RISE experiments that included juvenile offenders ($n=512$) discovered a strikingly consistent replication of the patterns Bennett (2008*) found with burglary and robbery offenders in London. While the mean frequency of arrests in the RISE 5-year baselines (about two arrests per year for crimes with personal victims in the highest-frequency trajectory, and less than one per year in the lowest) was far lower than in the London tests, RISE also showed a curvilinear pattern of RJC working better on high-rate and low-rate offenders than medium-rate offenders. Using an even longer follow-up period in Canberra than Bennett could use in

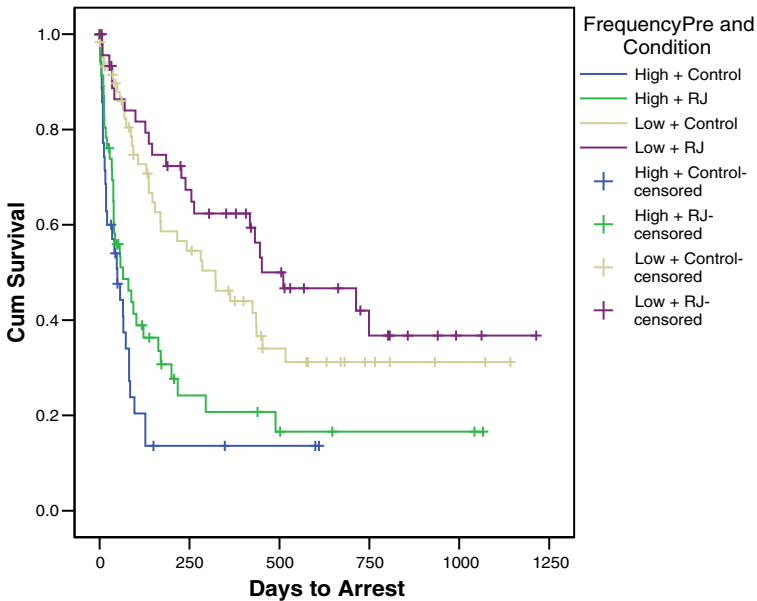


Fig. 1 Source: Bennett 2008*: 162

London (a 5-year follow-up after the 5-year baseline for all Canberra cases, for a total of 10 years of measurement), Woods used annual frequency of arrests of a specific kind (rather than time-to-failure for any new offense, as in London) as the outcome measure.

Woods (2009*) grouped all offenders in the three RISE experiments with juveniles into six trajectories of frequency of arrests for *crimes with personal victims only* (using trajectory analysis as described by Nagin 2005). His premise was that the RJC emphasis on empathy with victim suffering would be best tested by its impact on crimes against victims, as opposed to drug possession, drink-driving and other offenses without personal victims.

Woods then adjusted for the moderating effects of restorative justice with Aboriginal versus non-Aboriginal offenders, which led to his omitting all of the Aboriginal offenders from his final trajectory analysis, including two outlier cases that later analysis suggested to be driving overall findings about Aboriginals (Sherman et al. 2015a*, b*). Woods' decision in 2009 had the effect of reversing an initial (1 year after random assignment) increase in arrest frequency among highest-frequency offenders receiving RJCs (as Fig. 2 shows in the solid line rather than the dotted line controls in the same trajectory group). This procedure showed the biggest benefits of RJCs in reducing recidivism frequency among the most frequent offenders in the baseline period.

Conclusion #17: In three RISE tests and the robbery and burglary experiments in London, RJCs had the biggest effects on reducing recidivism on those offenders who had the highest rates of offending in the baseline period, and modest effects on very low-rate or first offenders, but was ineffective or criminogenic for those offenders with medium rates of offending in the baseline period.

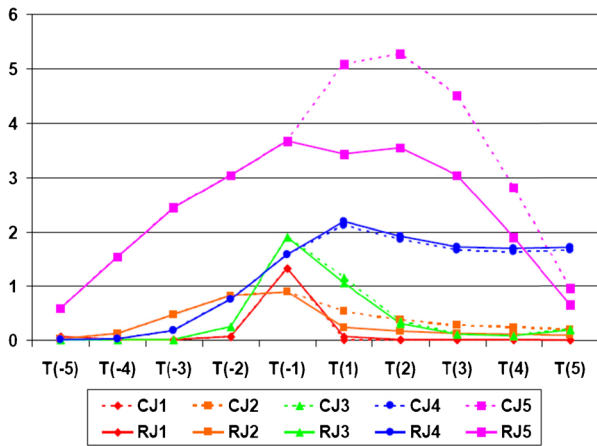


Fig. 2 Source: Woods 2009*: non-aboriginal subsample—impact of *RJ* vs. *CJ* on arrests for crimes with personal victims in three RISE Experiments (violence, property crime and shoplifting)

Repeat offending and offender multiple drug use

The link between drugs and crime is perhaps most hotly debated when discussing justice for drug-using offenders. The complexity of that debate runs into moderator effects on justice with offenders using different kinds of drugs one-at-a-time, or the difference between people using only one kind of drug vs. two or more kinds of illicit drugs simultaneously. Bennett (2008*: 202–204) used this discussion to examine any moderator effects of drug use patterns of the effects of RJC on time-to-failure. She found the London experiments offered a good opportunity. While 89 % of the London robbery and burglary offenders were reported to be using drugs at the time of arrest, only 53 % of burglars and 37 % of robbers were using both crack cocaine and heroin (combined $n = 152$). For those who did *not* use *both* crack and heroin, assignment to an RJC raised the mean days to first offense by 26 %, from 355 days to 447. But for offenders who *did* use *both* heroin and crack, assignment to an RJC backfired, by reducing their time to failure 29 %. The mean number of days at risk to first offense was 340 in the control group, but only 242 in the RJC group. The evidence thus supports this assessment:

Conclusion #18: London offenders who used both crack and heroin reoffended more quickly if they had been assigned to RJC than to controls, but offenders who did not use that combination of drugs reoffended more slowly if they were assigned to RJC than to controls.

Race and restorative justice

Early evidence in RISE suggested that RJC had been criminogenic for Aboriginal offenders (Strang and Sherman 2015*). Subsequent analyses have called this conclusion into question (Sherman et al. 2015a*, b*) and will be the subject of a detailed analysis in a future report.

Discussion: more work to be done

It seems unlikely that the 18 conclusions distilled in this review would have been produced in an ad hoc, one-RCT-at-a-time collection of experiments. The conclusions repeatedly draw on comparisons of answers to similar research questions across different kinds of offenses, offenders, and stages of the criminal process, as well as different countries. The external validity of the collective findings when analyzed in this fashion would seem to be far greater than what might be possible with 12 different experiments done by different research teams and organizations. That said, the addition of the independent evaluators in the UK experiments, combined with a standard approach to experimental design by the Jerry Lee Program, adds extra credibility to the external validity of the patterns (see Eisner 2009; Sherman and Strang 2009b). Given the frequent lack of any replication of policy experiments, with too many variations in practices being tested (and control groups compared to them) even when experiments are repeated, the Jerry Lee Program has clearly been different.

With this compilation of findings as an example, we are now able to make a stronger case in favor of governments and foundations obtaining greater benefits from a program of RCTs, rather than providing the same amount of funding for an ad hoc collection of experiments. Yet we must also ask whether we have made the most of the opportunity provided to us by a 12-RCT program. We can answer that question by reflecting on what else might be done with evidence from the Program, and specifically what we can aim to accomplish in the near-term.

There seems to be sound argument for three priorities: (1) we should publish more theoretically-focused articles or books that would feed the academic appetite for advancing theories, and not just facts, about crime and justice; (2) we should produce more highly specific manuals for practitioners, or “field guides” for how to create “suction” of criminal cases into RJs in different settings; and (3) we should push even harder to test RJs in more controversial areas, such as serious crimes, where our evidence shows that the benefits in harm reduction would be far greater for crime victims than where it is currently used.

But how does it work in theory?

One obvious way to get knowledge into practice is to make the knowledge more central academically, not just professionally. This is obvious because academics are the primary knowledge brokers on crime policy. While the professional or political demand for knowledge about justice innovations may not be great, the opportunities to supply knowledge may be heavily concentrated in the hands of university-based criminologists. These scholars not only advise the media and their local justice agencies on their opinions of what works. Academics also shape the views of tens of thousands of students who may go on to make and deliver justice policies.

Despite the 75 publications listed in the [Appendix](#), the Jerry Lee Program has arguably made little dent in academic thinking about justice innovations. Had at least some of the publications taken a more explicitly theoretical approach, there may have been more attention paid to restorative justice in undergraduate courses on the criminal justice processes. There might even have been more academically-initiated experiments

and research on RJs in a wider range of jurisdictions, offense types, and stages of the criminal justice process.

How do we know there has been little academic impact of the findings to date? One indicator is as simple as Google Scholar citation counts. Of the top ten publications listed when the words “Restorative Justice” are entered into Google Scholar, only three contain data from the Jerry Lee Program. Of those three, the highest citation count (1642 since a 2002 publication, or 130 Citations per year) is for the most theoretically elaborated interpretation of the experimental evidence (Braithwaite 2002). Other highly cited work is also more theoretical than the majority of the publications we have produced, which emphasize the empirical results over their theoretical meaning.

Why is it so important to use theory to gain academic attention and credibility? The answer is not limited to academics. The desire for understanding *why* something is true (Tilly 2006) is quite general, and may affect people’s willingness to believe that something really *is* true. Closely related to the desire to know why is a preference for *stories* over statistics, as the key funder of our Program, the radio broadcasting entrepreneur Jerry Lee of Philadelphia, has so often said. Stories about people provide a narrative that allows readers of any background to empathize with anyone—including offenders or victims who have been offered or denied RJs. A decade ago, we suggested the power of experimental ethnography, as a marriage of quantitative and qualitative methods, to address this appetite (Sherman and Strang 2004a). Yet, we have so far not produced a rigorously theoretical, let alone a qualitative–quantitative, analysis of our programmatic evidence in a mainstream peer-reviewed criminology or social science journal.

A field guide to getting criminal cases

At the opposite end of the continuum of theory to practice, we have failed to provide enough how-to-do-it instruction for practitioners. The need for such guidance is evident in every new initiative that is funded to provide restorative justice. Every such initiative of which we have heard has crashed against a wall of too few cases being offered for a program to be viable. Even the initiatives funded by the Home Office in 2001 that were not RCTs faced far greater difficulties than we did in generating cases that were dealt with by restorative justice.

We arguably have a lot of ‘good practice’ to share, at least in terms of implementation. Including our UK (non-controlled) Phase I practice cases, the Jerry Lee Program in 2001–2005 recruited over 1000 cases in which both offenders and victims agreed to meet (some 400 of which were randomly assigned to control groups). As far as we know, no other organization has ever produced 1000 cases in which full agreement was reached to conduct RJs. How we did it is something that can be spelled out, but it is usually too detailed for academic or scientific publications.

A case in point was recently suggested by the experience of the post-2013 legislative authorization of Judges adjourning cases for RJs prior to sentencing in Crown Court. That is exactly what we had tested in London in 2001–2005, obtaining some 500 cases of agreements by victims and offenders. Yet when Home Office funding was provided in 2014–15, the practitioners could hardly extract any cases from the Crown Court in which to conduct RJs (Collins 2015). Why was it so much harder to get cases in normal practice than in our tests?

The best explanation appears to be the decision of Judges supervising RJC in 2014 to diverge substantially from our practice in 2001–2005. They required that in order to conduct an RJC between guilty plea and sentence, the victim had to agree to do so even *before* the offender had pled guilty—which many of them do at the last minute. Not only did the RJ staff have zero time to ask the victims in the latecomer cases, they also could rarely assure victims that the offender was planning to plead guilty, nor could they say whether the offender was willing to meet with their victim. This system differed from what we tested in at least three respects: (1) we had been allowed time by Judges after each guilty plea to go *first* to the offenders, and only *second* to the victims, to seek consent for an RJC; (2) we had police officers, rather than “civilians,” approaching both offenders and victims for consent; and (3) we offered the assurance that the RJC itself would also be conducted by a police officer, which may have inspired some confidence in both offenders and victims that they would be protected from physical violence or other disorder by a police presence.

These details may seem petty, but they could also be the small things that make a big difference, the tipping points between getting cases or not getting cases. In justice experiments, the importance of conducting programs in *exactly the same administrative system* as they have been tested in RCTs is not widely understood. In contrast to medicine, where every tiny step of a medical procedure or pharmacological treatment is micro-managed, justice systems tend to be highly variable. There is no tradition in justice of worrying about little things making a difference, even though they might.

To be fair to the Judges in 2014, however, they could ask the Jerry Lee Program a very good question: “Why did you not write up the exact methods you used in successfully suctioning 1000 cases into RJC?” The answer is less important than the premise. The fact is that we did not spell out the procedures we used at the level of detail necessary for anyone to codify “best practice” for implementation. We did touch on it in a kind of field guide for youth justice practices (Sherman et al. 2008), but we did not produce field guides specific to different settings, such as Crown Courts. Nor did we pursue the issue of police versus civilians in their ability to recruit victims and offenders, which remains a key policy and funding issue in delivering RJC. Nor, in fact, did we offer to provide seminars to Crown Court Judges after our research results were analyzed, despite general invitations from individual judges to do so, another lacuna we regret.

To each according to their need

Perhaps the most serious critique of the Jerry Lee Program is that we have failed to convince policymakers that RJC are better used for serious cases and with chronic offenders than with minor crimes by juveniles and first offenders. Our unsystematic observation is that far more RJC are conducted with minor matters than with serious crimes and criminals. Our evidence shows that this is poor triage, giving RJC to people who have little need of it, and denying it to those whose need is greatest. If there is one conclusion that we should try to spread to a very wide audience, it is this one. How we can do that remains a question we cannot answer, except by the basic tools we use for all our work: grounded theory, trial and error, and systematic evidence.

It is not just the Jerry Lee Program that needs more knowledge about spreading knowledge effectively. It is all of experimental criminology, and science itself. This

article not only gives us a chance to reflect on how to put knowledge to work. It should give our readers the same opportunity, if only by thinking about how our Program could do better.

We close with one key plan for further research and analysis, driven in large part by the preceding discussion. The plan is to follow-up on the mortality differences between victims and offenders in the UK experiments, testing for any effects of RJC on life expectancy. Our evidence from 121 offenders under age 30 in one of the RISE tests is highly suggestive (Angel et al. 2013): while none of the 62 offenders randomly assigned (1995–2000) to the RJC group in the violence experiment had died by 2013, fully 10 % (6) of the 59 assigned to prosecution were dead (Fisher's Exact $P=.01$). In the UK, we can explore similar questions for victims with psychiatric evidence on PTSS. If we are able to find medical evidence that lower PTSS levels predict longer life span, we may well get more attention from governments, judges and police. We must be mindful of the responsibility we have to pursue this question, with the fully identified records of over 2000 people in our safekeeping. It may well be that RJC, like other criminal justice decisions (Sherman and Harris 2013, 2015), could be a matter of life and death.

Acknowledgments We wish to acknowledge the vision of John Braithwaite, who initiated this research program, and the financial support of the following funders: Australian Research Council, Australian Criminology Research Council, Australian Department of Health, Australian Department of Transport, US National Institute of Justice, Smith Richardson Foundation, Home Office for England and Wales, Esmee Fairbairn Foundation, Barrow Cadbury Trust, George Pine, University of Pennsylvania, and the Robert Wood Johnson Foundation. We also wish to thank the participating agencies that made the research possible, including the Australian Federal Police, Metropolitan Police Service, London Crown Courts, London Probation Service, Northumbria Police, Northumbria Magistrates' Courts, Northumbria Probation Service, Thames Valley Probation Service, HM Prison Bullingdon. We reserve our very special thanks to Jerry Lee and the Jerry Lee Foundation which made this RCT program possible.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Publications arising from the Jerry Lee Program of Randomized Trials in Restorative Justice 1995–2015

- Ahmed, E., Harris, N., Braithwaite, J., & Braithwaite, V. (2001). *Shame management through reintegration*. Melbourne: Cambridge University Press.
- Angel, C. M. (2005). *Crime victims meet their offenders: Testing the impact of restorative justice conferences on victims' post-traumatic stress symptoms*. Ph.D. Dissertation in Nursing and Criminology, University of Pennsylvania.
- Angel, C. M., Strang, H., Sherman, L. W., Bennett, S., Inkpen, N., Keane, A., & Richmond, T. (2008). *The impact of restorative justice on post-traumatic stress symptoms in male and female victims of serious crime*. Paper presented to the 15th World Congress of Criminology, Barcelona, Spain.
- Angel, C. M., Barnes, G. C., & Hyatt, J. With Strang H. & Sherman, L.W. (2013). *The effects of restorative justice on offender mortality, health, and welfare: A ten year follow-up*. Paper presented to the American Society of Criminology annual Meeting, Chicago, IL.
- Angel, C. M., Sherman, L. W., 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 randomised, controlled trial. *Journal of Experimental Criminology*, 10, 291–307.
- Barnes, G. C. (1999). *Procedural justice in two contexts: Testing the fairness of diversionary conferencing for intoxicated drivers*. PhD Dissertation, University of Maryland, College Park, MD.
- Barnes, G. C., Hyatt, J. M., Angel, C. M., Strang, H., & Sherman, L. W. (2015). Are restorative justice conferences more fair than criminal courts? Comparing levels of observed procedural justice in the reintegrative shaming experiments (RISE). *Criminal Justice Policy Review*, 26, 103–130.
- Bennett, S. (2008). *Criminal careers and restorative justice*. PhD Dissertation, University of Cambridge, Cambridge.
- Braithwaite, J. (2002). *Restorative justice and responsive regulation*. New York: Oxford University Press.
- Braithwaite, J. (2014). Evidence for restorative justice. *Vermont Bar Journal*, 40, 18–27.
- Braithwaite, J., & Braithwaite, V. (2001). Part I: Shame, shame management and regulation. In E. Ahmed, N. Harris, J. Braithwaite, & V. Braithwaite (Eds.), *Shame management through reintegration* (pp. 1–69). Cambridge: Cambridge University Press.
- Dignan, J., Atkinson, A., Atkinson, H., Howes, M., Johnstone, J., Robinson, & Sorsby, A. (2007). Staging restorative justice encounters against a criminal justice backdrop. A dramaturgical analysis. *Criminology and Criminal Justice*, 7, 5–32.
- Harris, N. (2000). *Shame and shaming: An empirical analysis*. PhD Dissertation, Australian National University, Canberra.
- Harris, N. (2001). Shaming and shame: Regulating drink driving. In E. Ahmed, N. Harris, J. Braithwaite & V. Braithwaite (Eds.), *Shame management through reintegration* (pp. 73–210). Cambridge: Cambridge University Press.
- Harris, N., & Burton, J. (1997). *The reliability of observed reintegrative shaming, shame, defiance and other key concepts in diversionary conferences*. RISE Working Paper No. 5. Canberra, Australia: Centre for Restorative Justice, Law Program, Research School of Social Science. retrieved from www.crim.cam.ac.uk
- Harris, N., & Burton, J. (1998). Testing the reliability of observational measures of reintegrative shaming at community accountability conferences and at court. *Australian and New Zealand Journal of Criminology*, 31, 230–241.
- Inkpen, N. (1999). *Reintegrative shaming through collective conscience building*. PhD dissertation, Australian National University, Canberra.
- Mugford, S., & Inkpen, N. (2011). *Reintegration ceremonies revisited: Experience and theory for advancing the argument*. Paper presented to the 3rd National Conference on Restorative Justice, Raleigh, NC. Retrieved from <http://www.slideshare.net/stephenkmugford/reintegration-ceremonies-revisited-2>.
- Robinson, G., & Shapland, J. (2008). Reducing recidivism: a task for restorative justice? *British Journal of Criminology*, 48(3), 337–358.
- Rossner, M. (2008a). Healing victims and offenders and reducing crime: a critical assessment of restorative justice practice and theory. *Sociology Compass*, 2(6), 1734–1749.
- Rossner, M. (2008b). *Why emotions work: Restorative justice, interaction ritual and the micro potential for emotional transformation*. PhD Dissertation, University of Pennsylvania, Philadelphia.
- Rossner, M. (2011a). Emotions and interaction ritual: a micro analysis of restorative justice. *British Journal of Criminology*, 51, 95–119.
- Rossner, M. (2011b). Reintegrative ritual: Restorative justice and micro-sociology. In S. Karstedt, I. Loader & H. Strang (Eds.), *Emotions, crime and justice* (pp. 169–192). Oñati International Series on Law and Society. Oxford, UK: Hart Publishing.
- Rossner, M. (2012). Emotion across the lifecourse: the case for restorative justice with adults. In J. Bolitho, J. Bruce, & G. Mason (Eds.), *Restorative justice: Adults and emerging practice* (pp. 214–230). Institute of Criminology series (30 201). Sydney, Australia: Sydney Institute of Criminology.
- Rossner, M. (2013). *Just emotions: Rituals of restorative justice*. Oxford: Oxford University Press.
- Rossner, M., & Tait, D. (2011). Contested emotions: adversarial rituals in non-adversarial justice procedures. *Monash University Law Review*, 37, 241–258.
- Shapland, J., Atkinson, A., Colledge, E., Dignan, J., Howes, M., Johnstone, J., Pennant, R., & Sorsby, A. (2004). *Implementing restorative justice schemes (Crime Reduction Programme): A report on the first year*. London: Home Office.
- Shapland, J., Atkinson, A., Atkinson, H., Colledge, E., Dignan, J., Howes, M., Johnstone, J., Robinson, G., & Sorsby, A. (2006a). Situating restorative justice within criminal justice. *Theoretical Criminology*, 10(4), 505–532.
- Shapland, J., Atkinson, A., Colledge, E., Dignan, J., Howes, M., Johnstone, J., Robinson, G., & Sorsby, A. (2006b). *Restorative justice in practice: The second report from the evaluation of three schemes*. Sheffield: University of Sheffield, Centre for Criminological Research.

- Shapland, J., Atkinson, A., Atkinson, H., Chapman, B., Dignan, J., Howes, M., Johnstone, J. Pennant, R., & Sorsby, A. (2007). Restorative justice: the views of victims and offenders. *Ministry of Justice Research Series*, 3(07).
- Shapland, J., Atkinson, A., Atkinson, H., Dignan, J., Edwards, L., Hibbert, J., Howes, M., Johnstone, J., Pennant, R., & Sorsby, A. (2008). *Does restorative justice affect reconviction? The fourth report from the evaluation of three schemes*. Sheffield: University of Sheffield, Centre for Criminological Research.
- Shapland, J., Robinson, G., & Sorsby, A. (2011). *Restorative justice in practice: Evaluating what works for victims and offenders*. Oxford: Taylor & Francis.
- Sherman, L. (2003). Reason for emotion: reinventing justice with theories, innovations, and research – the American Society of Criminology 2002 Presidential Address. *Criminology*, 41, 1–38.
- Sherman, L. (2014). Experiments in criminal sanctions: labeling, defiance and restorative justice. *Advances in Criminological Theory*, 18, 149–176.
- Sherman, L. W., & Barnes, G. (1997). *Restorative justice and offenders' respect for the law. RISE Working Paper No.3*. Canberra, Australia: Centre for Restorative Justice, Research School for Social Sciences, Australian National University. Retrieved from www.crim.cam.ac.uk
- Sherman, L., & Strang, H. (1997). *The right kind of shame for crime prevention. RISE Working Paper No. 1*. Canberra, Australia: Centre for Restorative Justice, Research School for Social Sciences, Australian National University. Retrieved from www.crim.cam.ac.uk.
- Sherman, L. W., & Strang, H. (1997). *Restorative justice and deterring crime. RISE Working Paper No. 4*. Canberra, Australia: Centre for Restorative Justice, Research School for Social Sciences, Australian National University. Retrieved from www.crim.cam.ac.uk.
- Sherman, L., & Strang, H. (2004a). *Restorative justice: What we know and how we know it. The Jerry Lee Program of Randomized Controlled Trials in Restorative Justice*. Philadelphia: Jerry Lee Center of Criminology, Department of Criminology, University of Pennsylvania.
- Sherman, L., & Strang, H. (2004b). Verdicts or inventions? Interpreting results from randomised controlled experiments in criminology. *American Behavioral Scientist*, 47, 575–607.
- Sherman, L. W., & Strang, H. (2004c). Experimental ethnography: the marriage of qualitative and quantitative research. *The Annals of the American Academy of Political and Social Science*, 595(1), 204–222.
- Sherman, L., & Strang, H. (2005). Evidence-based justice (correspondence). *The Lancet*, 365, 469–470.
- Sherman, L., & Strang, H. (2007). *Restorative justice: The evidence*. London: Smith Institute.
- Sherman, L., & Strang, H. (2009a). Crime and reconciliation: experimental criminology and the future of restorative justice. *Acta Criminologica*, 22, 1–14.
- Sherman, L., & Strang, H. (2009b). Testing for analysts' bias in crime prevention experiments: can we accept Eisner's one-tailed test? *Journal of Experimental Criminology*, 5, 185–200.
- Sherman, L., & Strang, H. (2010). Restorative justice as a psychological treatment: Healing victims and reintegrating offenders. In G. Towl & D. Crichton (Eds.), *Handbook of forensic psychology* (pp. 398–412). Amsterdam: Elsevier.
- Sherman, L., & Strang, H. (2011). Empathy for the devil: Nature and nurture in restorative justice. In S. Karstedt, I. Loader, & H. Strang (Eds.), *Emotions, crime and justice* (pp. 145–168). Oxford: Hart.
- Sherman, L., & Strang, H. (2012). Restorative justice as evidence-based sentencing. In J. Petersilia & K. Reitz (Eds.), *The Oxford handbook of sentencing and corrections* (pp. 215–243). New York: Oxford University Press.
- Sherman, L. W., Braithwaite, J., Strang, H., Barnes, G. C., Christie-Johnston, J., Smith, S., & Inkpen, N. (1997). *Experiments in restorative policing 1997: Reintegrative shaming of violence, drink driving & property crime: a randomised controlled trial*. Canberra: Australian National University, Research School of Social Sciences.
- Sherman, L. W., Strang, H., Barnes, G. C., Braithwaite, J., Inkpen, N., & Teh, M. (1998). *Experiments in restorative policing 1998: a progress report on the Canberra Reintegrative Shaming Experiments (RISE)*. Canberra: Australian National University, Research School of Social Science.
- Sherman, L.W., Strang H., & Woods, D.J. (2000). *Recidivism patterns in the Canberra Reintegrative Shaming Experiments (RISE)*. Canberra: Centre for Restorative Justice, Research School for Social Sciences, Australian National University. Retrieved from www.crim.cam.ac.uk.
- Sherman, L., Strang, H., & Woods, D. (2003). Captains of restorative justice: Experience, legitimacy and recidivism by type of offence. In E. G. M. Weitekamp & H. J. Kerner (Eds.), *Restorative justice in context: International practice and directions* (pp. 229–256). Devon: Willan.
- Sherman, L., Strang, H., Angel, C., Woods, D., Barnes, G., Bennett, S., Rossner, M., & Inkpen, N. (2005). Effects of face-to-face restorative justice on victims of crime in four randomised controlled trials. *Journal of Experimental Criminology*, 1, 367–395.
- Sherman, L., Strang, H., & Newbury-Birch, D. (2008). *Restorative justice*. London: Youth Justice Board.

- Sherman, L., Strang, H., Mayo-Wilson, E., Woods, D., & Ariel, B. (2015a). Are restorative justice conferences effective in reducing repeat offending? Findings from a Campbell systematic review. *Journal of Quantitative Criminology*, 31, 1–24.
- Sherman, L., Strang, H., & Slothower, M. (2015b). A long-term followup of criminal histories, by race, of offenders randomly assigned to court or restorative policing. Paper presented to the Stockholm Criminology Symposium, Stockholm, Sweden.
- Strang, H. (2000). *Victim participation in a restorative justice process: The Canberra Reintegrative Shaming Experiments*. PhD Dissertation, Australian National University, Canberra.
- Strang, H. (2001). Justice for victims of young offenders: The Centrality of Emotional Harm and Restoration. In A. Morris & G. Maxwell (Eds.), *Restorative conferencing for young offenders*. Oxford: Hart.
- Strang, H. (2002). *Repair or revenge? Victims and restorative justice*. Oxford: Oxford University Press.
- Strang, H. (2004). The threat to restorative justice posed by the merger with community justice: a paradigm muddle. *Contemporary Justice Review*, 7, 75–79.
- Strang, H. (2006). Higaisha Mondai to Shufukuteki Shiho (Victim Problems and Restorative Justice). In Y. Hosoi, H. Nishimura, S. Kashimuri & B. Tatsuno (Eds), *Shufukuteki Shiho Sougouteki Kenkyu (Restorative Justice: An Overview of Research)*(pp 69–83). Tokyo, Japan: Kazamashobo.
- Strang, H. (2010). Exploring the effects of restorative justice on crime victims for victims of conflict in transitional societies. In S. Shoham & P. Knepper (Eds.), *International handbook of victimology*. London: Taylor & Francis.
- Strang, H. (2011). *Ten year victim impact of restorative justice on victims in the Canberra RISE Experiments*. Paper presented to the American Society of Criminology Annual Meeting, Washington, DC.
- Strang, H. (2012a). Restorative justice and victims of crime: Evidence and possibilities. In H. Morosawa, J. Dussich, & G. F. Kirchhoff (Eds.), *Victimology and human security: New Horizons*. Nijmegen: Wolf Legal Publishers.
- Strang, H. (2012b). Conferencing and victims. In E. Zinsstag & I. Vanfraechem (Eds.), *Conferencing and restorative justice: International practices and perspectives*. Oxford: Oxford University Press.
- Strang, H. (2012c). Coalitions for a common purpose: managing relationships in experiments. *Journal of Experimental Criminology*, 8, 211–225.
- Strang, H., & Sherman, L. W. (1997). *The victim's perspective*. RISE Working Paper No. 2. Canberra, Australia: Centre for Restorative Justice, Research School for Social Sciences, Australian National University.
- Strang, H., & Sherman, L. W. (2003). Repairing the harm: victims and restorative justice. *Utah Law Review*, 1, 15–42.
- Strang, H., & Sherman, L. (2006). Restorative justice to reduce victimization? In B. Welsh & D. Farrington (Eds.), *Preventing crime: What works for children, offenders, victims and places*. Dordrecht: Springer.
- Strang, H., & Sherman, L. (2012). Experimental criminology and restorative justice: Principles of developing and testing innovations. In D. Gadd, S. Karstedt, & S. F. Messner (Eds.), *The Sage handbook on criminological research methods* (pp. 395–409). London: Sage.
- Strang, H., & Sherman, L. (2015). The morality of evidence: second annual restorative justice lecture. *Restorative Justice: An International Journal*, 3, 1–15.
- Strang, H., Barnes, G. C., Braithwaite, J., & Sherman, L. W. (1999). *Experiments in restorative policing: A progress report*. Law program, Research School for Social Sciences Australian National University (posted at www.crim.cam.ac.uk.)
- Strang, H., Sherman, L., Angel, C. M., Woods, D. J., Bennett, S., Newbury-Birch, D., & Inkpen, N. (2006). Victim evaluations of face-to-face restorative justice conferences: a quasi-experimental analysis. *Journal of Social Issues*, 62, 281–306.
- Strang, H., Sherman, L., Woods, D., & Barnes, G. (2011). *Experiments in restorative policing: Final report on the Canberra Reintegrative Shaming Experiments (RISE)*. Canberra, Australia: Law program, Research School for Social Sciences, Australian National University. Retrieved from www.crim.cam.ac.uk
- Strang, H., Sherman, L., Mayo-Wilson, E., Woods, D. J., & 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 Systematic Reviews* 2013(9). doi: 10.4073/csr.2013.10
- Tyler, T., Sherman, L., Strang, H., Barnes, G., & Woods, D. (2007). reintegrative shaming, procedural justice and recidivism: the engagement of offenders' psychological mechanisms in the Canberra RISE Drinking-and-Driving Experiment. *Law and Society Review*, 41, 553–586.
- Woods, D. J. (2003). *Self-reported vs. official offending in the Reintegrative Shaming Experiments (RISE)*. M.A. Thesis, Department of Criminology and Criminal Justice, University of Maryland.

Woods, D. J. (2009). *Unpacking the impact of restorative justice in the RISE experiments: facilitators, offenders and conference non-delivery*. PhD Dissertation, University of Pennsylvania, Philadelphia.

References

- Ariel, B., Farrar, W. A., & Sutherland, A. (2014). The effect of police body-worn cameras on use of force and citizens' complaints against the police: a randomized controlled trial. *Journal of Quantitative Criminology*, *31*, 509–535.
- Berlin, J. A., & Ghersi, D. (2005). Preventing publication bias: Registries and prospective meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Hoboken: Wiley.
- Boruch, R. F. (1997). *Randomized experiments for planning and evaluation: A practical guide* (Vol. 44). Thousand Oaks: Sage.
- Braithwaite, J. (1989). *Crime, shame and reintegration*. Cambridge: Cambridge University Press.
- Collins, R. (2004). *Interaction ritual chains*. Princeton: Princeton University Press.
- Collins, J. (2015). Remarks Made at a Forum of the Restorative Justice Council, London, October 9.
- Eisner, M. (2009). No effects in independent prevention trials: can we reject the cynical view? *Journal of Experimental Criminology*, *5*, 163–183.
- Farrington, D. P., & Petrosino, A. (2001). The Campbell collaboration crime and justice group. *The Annals of the American Academy of Political and Social Science*, *578*(1), 35–49.
- Gladwell, M. (2008). *Outliers: The story of success*. Boston: Little, Brown and Company.
- Jeong, S., McGarrell, E. F., & Hipple, N. K. (2012). Long-term impact of family group conferences on re-offending: the Indianapolis restorative justice experiment. *Journal of Experimental Criminology*, *8*(4), 369–385.
- Kubzansky, L., Koenen, K., Spiro, A., Vokonas, S., & Sparrow, D. (2007). Prospective study of posttraumatic stress disorder symptoms and coronary heart disease in the normative aging study. *Archives of General Psychiatry*, *64*(1), 109–116.
- MacDonald, J. (2015). See <http://proactive-resolutions.com/about/team/> downloaded 31 October.
- McGarrell, E. F., & Hipple, N. K. (2007). Family group conferencing and re-offending among first-time juvenile offenders: the Indianapolis experiment. *Justice Quarterly*, *24*(2), 221–246.
- Nagin, D. (2005). *Group-based modeling of development*. Cambridge: Harvard University Press.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, *349* aac4716, 1–8. doi: [10.1126/science.aac4716](https://doi.org/10.1126/science.aac4716)
- Sherman, L. W. (2007). The power few: experimental criminology and the reduction of harm. *Journal of Experimental Criminology*, *3*(4), 299–321.
- Sherman, L. W. (2013). The rise of evidence-based policing: targeting, testing, and tracking. *Crime and Justice*, *42*(1), 377–451.
- Sherman, L. W. (2015). A tipping point for “Totally Evidenced Policing”: ten ideas for building an evidence-based police agency. *International Criminal Justice Review*, *25*, 11–29.
- Sherman, L. W., & Harris, H. M. (2013). Increased homicide victimization of suspects arrested for domestic assault: a 23-year follow-up of the Milwaukee Domestic Violence Experiment (MilDVE). *Journal of Experimental Criminology*, *9*, 491–514.
- Sherman, L. W., & Harris, H. M. (2015). Increased death rates of domestic violence victims from arresting vs. warning suspects in the Milwaukee Domestic Violence Experiment (MilDVE). *Journal of Experimental Criminology*, *11*, 1–20.
- Sherman, L. W., Schmidt, J. D., & Rogan, D. P. (1992). *Policing domestic violence: Experiments and dilemmas*. New York: Free Press.
- Strang, H. (2015). Email communication 17 October 2015 with Jon Collins, Chief Executive Officer, (UK) Restorative Justice Council.
- Tilly, C. (2006). *Why?: What happens when people give reasons... why*. Princeton: Princeton University Press.
- Weinberger, M., Oddone, E. Z., Henderson, W. G., Smith, D. M., Huey, J., Giobbie-Hurder, A., & Feussner, J. R. (2001). Multisite randomized controlled trials in health services research: scientific challenges and operational issues. *Medical Care*, *39*(6), 627–634.
- Weisburd, D., Petrosino, A., & Mason, G. (1993). Design sensitivity in criminal justice experiments. *Crime and Justice*, *17*, 337–379.

Lawrence W. Sherman is Wolfson Professor of Criminology and Director of the Jerry Lee Centre of Experimental Criminology at the Institute of Criminology, University of Cambridge, as well as Distinguished University Professor at the Department of Criminology and Criminal Justice, University of Maryland in College Park. His most recent article in this Journal, with Heather Harris, was “Increased death rates of domestic violence victims from arresting vs. warning suspects...” (2015).

Heather Strang is Director of the Police Executive Programme and Director of Research of the Jerry Lee Centre for Experimental Criminology at the Institute of Criminology, University of Cambridge. Author of “Revenge or Repair: Victims and Restorative Justice” (2002), her most recent article in this Journal was “Coalitions for a common purpose: managing relationships in experiments” (2012).

Geoffrey C. Barnes is Research Assistant Professor at the Jerry Lee Center of Criminology at the University of Pennsylvania, and a Visiting Scholar at the Police Executive Programme, Institute of Criminology, University of Cambridge. His most recent article in this Journal, with others, was “Low-intensity community supervision for low-risk offenders: a randomized, controlled trial” (2010).

Daniel J. Woods is Senior Research Associate at the Police Foundation in Washington, DC, and a former Jerry Lee Scholar at the University of Pennsylvania. His most recent contribution to this Journal, with Bruce Taylor and Christopher Koper, was “A randomized controlled trial of different policing strategies at hot spots of violent crime” (2011).

Sarah Bennett is a Senior Research Fellow in the School of Social Science, Faculty of Humanities and Social Sciences at the University of Queensland. A former program manager in the London Restorative Justice Experiments at the Metropolitan Police, her most recent contribution to this journal (with Lorraine Mazerolle and others) was “Procedural justice and police legitimacy: a systematic review of the research evidence” (2013).

Nova Inkpen is Justice Reinvestment Manager in the Australian Capital Territories government’s Justice and Community Safety Directorate. Her most recent contribution to this Journal, with Caroline Angel and others, was “Short-term effects of restorative justice conferences on post-traumatic stress symptoms among robbery and burglary victims” (2015).

Dorothy Newbury-Birch is Professor of Alcohol and Public Health Research at the University of Teesside, UK. Her most recent experiment was published (with others) in *Alcohol and Alcoholism* (2014) as “Screening and brief interventions for hazardous and harmful alcohol use for offenders: a pragmatic multicentre cluster randomised controlled trial.”

Meredith Rossner is Assistant Professor of Criminology at the London School of Economics. Her book, “Just Emotions: Rituals of Restorative Justice”, was published in 2013 by Oxford University Press.

Caroline M. Angel is a Visiting Scholar at the University of Cambridge and Research Director and Central NJ Team Captain at Team Red, White & Blue. Her last contribution to this Journal, with others, was “Short-term effects of restorative justice conferences on post-traumatic stress symptoms among robbery and burglary victims” (2015).

Malcolm Mearns is Chief Executive of DATACOL in Canberra, ACT, Australia, where he managed all of the survey data collected for the RISE project and its long-term followup.

Molly Slothower is a Junior Fellow at the Cambridge Centre for Evidence-Based Policing Ltd. and a Ph.D. student in the Department of Criminology and Criminal Justice at the University of Maryland, College Park. She served as field manager for the West Midlands Police Turning Point experiment in Birmingham for two years, and the project manager for the Blueprints for Gang Violence Prevention experiment in Philadelphia. She continues to analyze data from both that project and the Canberra RISE criminal history followup.