

Are Randomised Controlled Trials Really the 'Gold Standard' in Restorative Justice Research?

ARE RANDOMISED CONTROLLED TRIALS REALLY THE 'GOLD STANDARD' IN RESTORATIVE JUSTICE RESEARCH?

Aidan Wilcox with Carolyn Hoyle and Richard Young, University of Oxford Centre for Criminological Research.

Abstract

The rise of the evidence-based approach has influenced how criminological research is conducted in Britain. Funding bodies increasingly specify a quantitative research design as the most appropriate way to determine 'what works'. Indeed, there is a growing tendency to regard the randomised control trial (RCT) as the 'gold standard' for such studies. In this article, restorative justice research is used as an example of the pitfalls of relying on a purely quantitative paradigm. By ignoring issues of programme integrity and context, RCTs often produce contradictory or inconclusive results. The article concludes that the artificial divide often imposed on researchers between qualitative and empirical approaches is unhelpful, and that research into restorative justice (and other areas) would gain from the integration of an in-depth, qualitative component within the experimental approach.

Introduction

This article considers two differing approaches to criminal justice evaluation, using restorative justice research to illustrate the arguments developed. The first approach is the quantitative, with its emphasis on randomised controlled trials (RCTs), which has been encouraged by the contemporary evidence-based approach to policy. To their advocates, RCTs hold out the promise of delivering definitive judgements as to 'what works'. In practice, however, RCTs remain not only rare but are frequently dogged by methodological weaknesses, which limit their contribution to the policy process. Even where RCTs have been implemented as planned, they are often unable to provide unambiguous, policy-relevant answers. The lack of attention to process means that the mechanisms through which outcomes are achieved may be overlooked and that the results of apparently similar tests of an intervention are sometimes contradictory. In contrast, the strength of the qualitative method lies not in determining causality, but in investigating

Aidan Wilcox, Carolyn Hoyle and Richard Young

the mechanisms through which interventions may work and the reasons why they appear to work for some and not others. A combination of the strengths of both approaches is needed if policy is really to be evidence-based.

Evidence Based Policy and Research

The relationship between social research and policy has fluctuated over the years. For example, the pessimism of the 'nothing works' mood of the 1970s (e.g. Martinson, 1974) led to a decline in the influence of research on policy. However, by the mid 1980s academic support for a revived 'what works' agenda began to emerge. One example was the probation research of McGuire and Priestley (1985), which suggested that some forms of interventions with offenders were effective. A decade later, the government explicitly adopted an evidence-based, or 'what works', approach to policy-making. In the criminal justice field, for example, the Home Office published a survey of 'what works' research (Vennard et al., 1997) with the aim of encouraging an evidence-based approach to probation practice.

The publication of more encouraging research findings was not the only reason for the increasing interest in, and support for, an evidence-based approach to policy. The election of the Labour government in 1997 signalled a shift away from ideological managerialism, with its emphasis on quasi-markets, targets and cost efficiency (Richards and Smith, 2002) towards a more pragmatic concern with 'what works'. In practice, an evidence-based approach is taken to mean that policy initiatives are to be supported by research evidence and that initiatives introduced on a trial basis are to be evaluated as rigorously as possible (Plewis, 2000). Government support has been translated into increases in research funding and the establishment of academic centres with the aim of collating, assessing and disseminating research findings (Wilcox, 2003).

Randomised Control Trials

Introduction

One of the consequences for criminological research generally (and research into restorative justice in particular) of the evidence-based, or 'what works', approach has been to promote the quantitative method of research. The Home Office, as a primary funder of external criminological research, has encouraged this trend, and it has been noted that much Home Office research is narrowly focused and atheoretical, with an emphasis on short term, empirical, quantitative methodology (Sebba, 2001:8). This more 'administrative' criminology has been concerned with determining 'what works' mainly through quantitative large-scale studies in which the outcome of interest (to the funder at least) is reconviction. The 'gold standard' of the quantitative approach is considered to be the randomised control trial (RCT), in which individuals are allocated at random to either the treatment or control condition. Ethical and practical difficulties mean that rigorous RCTs are still quite rare in criminal justice research: a draft review of RCTs of restorative interventions for the Campbell collaboration found just seven examples of completed RCTs, with a further eight ongoing (Strang and Sherman, 2003).

Are Randomised Controlled Trials Really the 'Gold Standard' in Restorative Justice Research?

The great strength of RCTs is claimed to be their ability to test causal hypotheses. When trying to test such a hypothesis (such as 'does a restorative conference reduce reconviction amongst violent offenders compared to a court appearance?') it is vitally important to be able to rule out alternative explanations for any observed differences in outcome between treatment and control groups. It is generally accepted that the best way to do this is through random assignment of individuals to treatment and control groups. Provided there are sufficient numbers in both groups, random assignment ensures that *on average* individuals in the treatment group will be similar to individuals in the control group *before* treatment is applied and therefore one can be more confident that any observed change is due to the treatment (Cook and Campbell, 1979). The next section of the article will consider some general criticisms of RCTs before looking in more detail at one example.

Implementation Issues

Although the RCT is potentially a strong research design, in practice implementation failures often weaken the causal inference which can be made (Lösel, 2001). For example, where attrition (drop outs, or loss of individuals from their initial allocation to the treatment or control group) is high there is the risk that those remaining in the groups are no longer representative of the group as a whole, and therefore one is no longer comparing the effect of the intervention to that of the control condition. The smaller the expected effect of the intervention, the greater the problem that attrition poses for the researcher. Similarly, where the attrition rates differ between the treatment and control groups, the interpretation of results is made difficult, since the reasons for drop out may differ, and one cannot be sure that the individuals remaining in the treatment group are the same, on average, as those remaining in the control group (Cook and Campbell, 1979). It may be the case, for example, that persistent offenders are more likely to fail to attend a restorative conference than a court hearing. Those remaining in the intervention group would therefore have, on average, a lower risk of reconviction than those remaining in the control group and it would be misleading to interpret a difference in reconviction rates between those remaining in the groups as resulting from the effect of the restorative conference. Although there is no universal agreement as to what level of attrition would be fatally damaging for a study, Strang and Sherman (2003) suggest that 'major' attrition be defined as 25% or more (although this was later raised to 34%: Sherman and Strang, 2003). In fact, the attrition rates in the RCTs they surveyed ranged from 33% to 68% (Strang and Sherman, 2003).

Many RCTs also suffer from small sample sizes, which means that the individuals in the two groups may not, on the whole, be similar. Low sample sizes also make it more difficult to establish the statistical significance of any results, especially where the expected difference in outcome (e.g. reconviction) is likely to be small. In four of the seven RCTs analysed by Strang and Sherman, there were fewer than 100 cases each in the comparison and intervention groups (Strang and Sherman, 2003). It was not surprising, therefore, that few of the comparisons between treatment and control groups reached the level of statistical significance.

Aidan Wilcox, Carolyn Hoyle and Richard Young

The Medical Comparison

Problems of attrition and low sample size can affect RCTs in any discipline, and point to the need to ensure that the evaluation is well designed and implemented as planned. However, specific problems can arise from the use of RCTs in criminal justice. Randomised control trials originated in medical research, and it is important to note the differences in how RCTs can be used in medical settings as opposed to social science ones. Chalmers et al (1981) discuss the use of RCTs in medicine and argue that the most important determination of the methodological quality of an RCT is the blinding procedure; that is to say, disguising from both those *administering* an intervention (doctor/facilitator) and those *receiving* it (patient/offender), which of the treatment or control conditions they are in or administering. The reason for this is that knowledge of the assignment may affect: the reaction of an offender to an intervention; the behaviour of the person providing it; or the assessment of the outcome. This is particularly important where the treatment, be it a medicine or a criminal justice intervention, is considered to be subjectively different (e.g. more/less effective, painful, intrusive) from the control condition, as it may lead to drop out or non-compliance amongst the group receiving the less favourable treatment. It has consistently been shown in medical research that where treatment is inadequately concealed the study is more likely to show a positive effect (Schultz et al., 1995).

Unlike medical research, however, it is rarely possible to use blinding procedures to disguise the 'treatment' in criminal justice settings (for example an offender will recognise the difference between going to court and participating in a restorative meeting). Nor is it possible to disguise from those administering an intervention which intervention they are providing. It is an often overlooked point that the consequence of this is that RCTs, as practised in criminal justice research, would be considered to be methodologically weak by medical criteria and would not normally be included in any review of the research. For example, the rating system proposed by Chalmers and colleagues awarded 30 of the 100 points available for rating the methodological quality of RCTs to the blinding procedures (Chalmers et al., 1981). Thus even the best conducted criminological RCT is a tarnished 'gold standard' by comparison. Indeed it is questionable whether the evaluations labelled as RCTs in criminology should really be defined as such, given the lack of blinding procedures (Bottomley, 2002). This inability to disguise treatment and the fact that social science experiments deal with reflexive human beings mean that evaluators have to think about and try to test other potential explanations for observed differences in outcomes.

Some Problems of Validity

In this part of the article, an example of an RCT in restorative justice will be used to highlight some of the issues which may arise from the lack of blinding and inattention to detail about processes and mechanisms. The RISE (reintegrative shaming experiments) study has been chosen because it is one of the few examples of a well-conducted RCT in restorative justice research. The RISE study looked at the effect of police-run restorative conferences as an alternative to court processing on outcomes including reconviction (Strang et al., 1999). Offenders were allocated at random to either the court or restorative

Are Randomised Controlled Trials Really the 'Gold Standard' in Restorative Justice Research?

conference. Careful measurement was made of differences between the two groups in terms of attitudes, experiences and outcomes. The study was implemented as intended: it did not suffer from small sample sizes; the randomisation process was largely successful in ensuring that participants in the conference and court groups were similar in terms of the main factors likely to influence reconviction (thus ruling out selection effects) and; attrition from both groups was relatively low. However, there are potential threats to validity which even a well designed and implemented experiment may face. Some of these threats, as they might apply to the RISE study, are described below (the relative importance of different threats to validity will, of course, depend on the nature of the research setting and outcome of interest).

The first of these relates to the interaction of different treatments (Cook and Campbell, 1979). In some criminal justice initiatives offenders may receive more than just the intervention under investigation. For example, offenders may receive additional support *after* a conference which did not specifically form part of the restorative intervention (such as mentoring, or help with education or training). It is difficult to establish causality if individuals experience more than one treatment. In the RISE study, the fact that those who attended a conference did not receive criminal records (Strang et al., 1999) might be considered to be a treatment in itself, and therefore a possible alternative explanation for the results. For example, the fact of not being officially labelled an offender may have lessened individuals' propensity to offend, quite independent of the restorative nature of the intervention (e.g. Becker, 1963).

Secondly, researchers sometimes fail to consider factors beyond the treatment or control conditions which might in fact have a greater impact on the results. For example, one of the findings of the RISE study was that participants felt that court was less fair than conference (Strang et al., 1999). But this might not be because conferences were inherently fairer, but may have been because in court the defendants would have had the chance to consult with a defence lawyer who may have raised their awareness about fairness – about what they should be able to expect of the process and their rights within it (Young, 2001). In a restorative conference there are typically no lawyers present to explain these matters. A similar point is that victims who attended a restorative conference might have been more satisfied simply because someone had taken notice of them, talked to them about their experience, and had given them information, rather than because of the restorative intervention per se. So it might be that Victim Support, carrying out these basic services in the courts, could bring about the same level of satisfaction.

The aim of many criminal justice interventions is to reduce reoffending. However, since offending cannot directly be measured, it is usual to measure reconviction as a proxy for reoffending. In so doing, it is assumed that changes in the level of reoffending will be mirrored by changes in the reconviction rate. However it is possible that reconviction rates will decrease (or increase), even while the true level of offending remains the same. If this were the case, it would provide an alternative explanation for observed differences

Aidan Wilcox, Carolyn Hoyle and Richard Young

between reconviction rates for court and conference groups. Braithwaite (2002) argues that, in principle, restorative justice should bring communities together and result in greater monitoring of offenders (the point being that restorative justice seeks to engage informal social control). This could in practice lead to increased reporting of further offending (because of increased monitoring and ownership of problem – doing something about it rather than ignoring it) or it could lead to the opposite – families and the community shielding the offender from further reporting because they are committed to familial/community reintegrative shaming. However, without research on the significant interactions between young offenders and their supporters for months after conferences it is not possible to tell which, if either, of these mechanisms is at work, to what extent, and with what degree of uniformity between different groups of offenders.

Furthermore, insofar as they have discretion over these matters, reactions of criminal justice officials to reoffending may differ according to the initial assignment of the offender to court or conference, for example they may be more likely to press for prosecution in cases where it is thought that the offender has not learnt from the restorative approach. In other words 'he had his chance and he blew it'. Similarly, it is possible that those who were assigned to court as a result of the RCT may be looked upon more sympathetically – 'she wasn't given a chance; let's give her one now by diverting the case from court'. In drawing comparisons between the original treatment and control groups, reconviction rates would then provide a particularly unreliable guide to the underlying offending rates. In our particular example any beneficial impact of restorative justice on reoffending may be completely obscured.

As mentioned above, the fact that individuals know which treatment condition they are assigned to can affect their reaction to the treatment, and one example of this is resentful demoralisation (Cook and Campbell, 1979). For those who are receiving subjectively less desirable treatments (such as going to court rather than to a conference) there may be a tendency for those in the control group to react with resentment and experience feelings of demoralisation which could lead to a spurious difference emerging between the two groups, at least in terms of satisfaction rates. If one considers Tyler's thesis which investigates why people obey the law (i.e. that the perceived legitimacy of the law and its enforcement is an important explanatory factor: Tyler, 1990), one can see that this might result in a greater propensity amongst the demoralised to reoffend. This is the point made by Killias and colleagues, who called it the 'bitterness effect'. In an experiment investigating the comparative effectiveness of community service versus short prison sentences, they found that those who were randomly assigned to prison rather than the community tended to experience feelings of embitterment. They argued that researchers should consider this as a potential explanation for observed differences and therefore attempt to measure it through interviews and estimate its impact (Killias et al., 2000).

One criticism which is often levelled at the quantitative approach, including RCTs, is that insufficient detail is given either to the nature of the intervention being researched or the extent to which it has been implemented as intended. As Marshall and Merry (1990)

Are Randomised Controlled Trials Really the 'Gold Standard' in Restorative Justice Research?

note, naming something is not the same as describing it. Restorative justice practices such as conferencing and mediation cover a wide range of diverse activities; different practices based on different theories and principles. The empirical focus of evidence-based policy risks focusing on the measurement of quantity over quality, because it is easier to do so and more likely to provide seemingly positive results (Crawford, 2001). Quantitative studies have been criticised for ignoring less tangible relationships between agencies and individuals which contribute to programme integrity and effectiveness (Cherney, 2002:54).

The Importance of Context

Scientific realists (e.g. Pawson and Tilley, 1994) argue that whilst RCTs may be able to demonstrate that an intervention was associated with a particular outcome, they often pay insufficient attention to the intervention itself, the contexts in which it operates and the ways through which it may work. It is not surprising, therefore, that RCTs often generate inconclusive or ambiguous results. The review by Strang and Sherman concluded that there was 'little evidence for the potential benefits of restorative justice in reducing reoffending for all offence types' and went on to suggest that this may be because effects for particular subgroups were being hidden, and that further RCTs were needed (Strang and Sherman, 2003: 29). Perhaps different types of RCTs are needed. While current RCTs may show that what was labelled as restorative justice had an effect on reconviction, they cannot necessarily say that this was *because* it was restorative justice (or even that it was restorative justice), why such an effect occurred or why it worked for some people and not others. Yet it is these latter questions which are of most interest to conscientious policy makers.

Incorporating a Qualitative Component

The above discussion points to the need for more in depth research if conclusions are to be drawn about which aspects of an intervention are associated with positive or negative outcomes. The next section uses the study carried out by Hoyle, Young and Hill (2002) in Thames Valley to highlight the strengths (and weaknesses) of this approach. It is not proposed that qualitative research alone could answer the kinds of questions posed above, as to why an intervention works and for whom. Rather, the study by Hoyle et al is used to illustrate briefly some of the qualitative methods which could be used in research and the kinds of issues which they can help to unravel.

The authors of the study deliberately adopted a qualitative approach in order to throw light on the mechanisms through which outcomes (such as on offender behaviour or victim satisfaction) were achieved. In doing so, it was hoped to complement the findings of RCTs such as the RISE study. The underlying premise of the Thames Valley research was that many criminal justice programmes 'fail' not because of the weakness of the underlying ideas but because of poor implementation. One possible reason for the 1970s conclusion that 'nothing works' was that little had been properly tried, still less properly evaluated (Young and Hoyle, 2000). Braithwaite notes that potential benefits of restorative justice interventions are often lost due to poor implementation, and a lack of

Aidan Wilcox, Carolyn Hoyle and Richard Young

'quality assurance' in many programmes (Braithwaite, 2002). Partly as a response to such concerns, it was decided that a formal evaluation of the impact of the Thames Valley initiative should only be attempted following a period of action-research in which the researchers would help the police implement its model as planned.

The research aims of the study were therefore not only to explore the relationships between restorative processes and outcomes, but also to measure the anticipated gap between restorative theory and policing practice, understand its causes and effects, and test whether it was possible for the police to close this gap once any failings in implementing the model had been documented through research. This would also allow the impact of 'police-led restorative justice', as opposed to police-led cautioning sessions labelled as such, to be tested.

The action-research methodology comprised a 'before and after' component. Through systematic 'before' measurements, including interviews with stakeholders and observation of conferences, the researchers made a number of recommendations (Young and Hoyle, 1999) which were accepted by Thames Valley Police, designed to close (or at least narrow) the gap detected between the programme's protocols and the behaviour of the facilitators. Following implementation of the recommendations, further qualitative and quantitative data were collected in the full evaluation, which involved observing and tape-recording 79 restorative conferences, tape-recording follow-up interviews with the great majority of the participants, resulting in over 600 interviews, and interviewing a sample of those who were affected by the offences but did not attend the cautioning sessions.

One of the most important methods used in the research was the observation and tape recording of all 79 restorative conferences within the sample. This allowed the researchers to examine the way in which particular contexts and mechanisms lead to outcomes at the individual level in a way in which purely quantitative studies are not able to do (Pawson and Tilley, 1994). For example, as a result of observation and through qualitative analysis of conference transcripts, the authors were able to arrange them into three categories of 'most', 'mid', and 'least' restorative by way of assessing quality of facilitation and adherence to the model. They could also assess how different aspects of facilitation impacted on, amongst other things, participants' views of fairness and changes in offenders' attitudes to victims and their own offending careers.

One criticism which could be levelled at this study is that the relatively small numbers involved do not permit statistically significant conclusions to be drawn. For example, one of the findings was that one in five offenders tended towards desistance at least in part due to the experience of restorative justice. While this conclusion was based on interviews, self-reports and reconviction data, the authors recognised that it could only be seen as suggestive without a larger-scale, comparative resanctioning study to test the validity of the finding (Hoyle, Young and Hill, 2002). Furthermore, since the authors were reliant on the police to notify them when conferences were being arranged, it is possible that some

Are Randomised Controlled Trials Really the 'Gold Standard' in Restorative Justice Research?

of the cases were 'cherry-picked', and thus not representative of cautions as a whole in the Thames Valley (although the problem of 'cherry-picking' is not unique to qualitative research and could equally occur in an RCT). However, the authors were able to demonstrate that the cautions they observed were representative of the total population of cautions administered in the three areas under evaluation during the sample period (Hoyle et al., 2002: 76).

Importantly, the study did not involve a comparison group of offenders receiving traditional cautions. In research which adopts a predominantly qualitative approach one is often unable to compare like with like, or even to make a comparison at all, and so there is no scope for making definitive causal statements. This did not mean that Hoyle et al were unable to assess the impact of the intervention on participants' perceptions of fairness, for example, but it did mean that the findings in relation to desistance could only be compared to results from other studies of old style cautions in other police areas.

Conclusion

The current evidence-based climate and concern with 'what works' has led to a preoccupation with quantitative, experimental approaches to evaluation which, as noted above, often produce contradictory findings and fail to explain how the same intervention can work in different ways with different effects according to the variations in the circumstances of those involved (Tilley, 2001:85). The qualitative approach has the advantage of being able to prompt interrogation and inquiry into the contexts and mechanisms associated with outcomes which can then be more fully tested through the experimental design of an RCT. *Within* the experimental design, incorporation of an in-depth approach can help researchers to identify, attempt to measure and deal with alternative explanations when the causal relationship is being tested.

Solesbury has argued that the 'what works' question is too simplistic, and that what policy makers need to know is what works for whom and under what circumstances. To answer this more complicated question, there is a need, he says, for research which is not just evaluative, but also 'descriptive, analytical, diagnostic, theoretical and prescriptive' (Solesbury, 2001: 8). To put this another way, whilst the recent research in restorative justice has largely been characterised by *either* qualitative or quantitative approaches, we would agree with Roger Hood's view that the quantitative and qualitative should not be regarded 'as a division of labour, for one cannot be done convincingly without the other' (Hood, 2002: 159). Such a combined approach would lead to improved reliability of findings, but would depend upon increased investment from the government and other funding bodies, since it would require both large samples and the resources to study the sample and the intervention in depth. The consequence of this may be that fewer studies are funded but the results of those studies would be less ambiguous and more policy-relevant. The discussion of the potential methodological weaknesses of RCTs has shown the risk that the policy implications flowing from such research may be inconsistent or incorrect. If reliance continues to be placed solely on the experimental approach in the misguided belief that it represents the 'gold standard', there is a danger that confidence in

Aidan Wilcox, Carolyn Hoyle and Richard Young

the value of criminological research will be (once again) seriously undermined (Bottomley, 2002).

References

- Becker, H. (1983) *Outsiders*. New York: Free Press.
- Bottomley, K. (2002) 'Editor's Tailpiece', *British Society of Criminology Newsletter* 46: 12.
- Braithwaite, J. (2002) *Restorative Justice and Responsive Regulation*. Oxford: Oxford University Press.
- Chalmers, T. C., Smith, H., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., and Ambroz, A. (1981) 'A Method for Assessing the Quality of a Randomized Control Trial', *Controlled Clinical Trials* 2: 31-49.
- Cherney, A. (2002) 'Beyond Technicism: Broadening the 'What Works' Paradigm in Crime Prevention', *Crime Prevention and Community Safety: An International Journal* 4(3): 49-59.
- Cook, T. D. and Campbell, D. T. (1979) *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.
- Crawford, A. (2001) 'Joined up but Fragmented: Contradiction, Ambiguity and Ambivalence at the Heart of New Labour's 'Third Way' ', in R. Matthews and J. Pitts (eds.) *Crime, Disorder and Community Safety: A New Agenda?* London: Routledge.
- Hood, R. (2002) 'Criminology and Penal Policy: The Vital Role of Empirical Research' in A. Bottoms and M. Tonry (eds.) *Ideology, Crime and Criminal Justice*. Cullompton: Willan.
- Hoyle, C., Young, R. and Hill, R. (2002) *Proceed with Caution: An Evaluation of the Thames Valley Police Initiative in Restorative Cautioning*. York: Joseph Rowntree Foundation.
- Killias, M., Aebi, M. and Ribeaud, D. (2000) 'Does Community Service Rehabilitate better than Short term Imprisonment? Results of a Controlled Experiment', *The Howard Journal*, 39(1): 40-57.
- Lösel, F. (2001) 'Evaluating the Effectiveness of Correctional Programs: Bridging the Gap Between Research and Practice', in G. Bernfeld, D. P. Farrington and A. W. Leschied (eds.) *Offender Rehabilitation in Practice*. London: Wiley and Sons.
- Marshall, T. and Merry, S. (1990) *Crime and Accountability: Victim/Offender Mediation in Practice*. London: HMSO.
- Martinson, R. (1974) 'What Works? - Questions and Answers about Prison Reform', *The Public Interest* 10: 22-54.
- McGuire, J. and Priestley, P. (1985) *Offending Behaviour: Skills and Stratagems for Going Straight*. London: Batsford.
- Pawson, R. and Tilley, N. (1994) 'What Works in Evaluation Research?', *British Journal of Criminology* 34(3): 291-306.
- Plewis, I. (2000) 'Educational Inequalities and Education Action Zones', in C. Pantazis and D. Gordon (eds.) *Tackling Inequalities: Where We Are Now and What Can be Done*. Bristol: Policy Press.
- Richards, D. and Smith, M. J. (2002) *Governance and Public Policy in the United Kingdom*. Oxford: Oxford University Press.
- Schulz K. F., Chalmers I., Hayes R. J. and Altman, D. G. (1995) 'Empirical Evidence of Bias: Dimensions of Methodological Quality Associated with Estimates of Treatment Effects in Controlled Trials', *Journal of the American Medical Association* 273: 408-412.
- Sebba, L. (2001) 'On the Relationship Between Criminological Research and Policy: The Case of Victims', *Criminal Justice* 1(1): 27-58.
- Sherman, L. and Strang, H. (2003) 'Verdicts or Inventions? Interpreting Results from Randomized Controlled Experiments in Criminology', *American Behavioral Scientist* 47(5): 575-607.
- Solesbury, W. (2001) *Evidence Based Policy: Whence it Came and Where it's Going*. Working Paper 1. London: ERSC UK Centre for Evidence Based Policy and Practice. Available online at: <http://www.evidencenetwork.org/Documents/wp1.pdf> (12 January 2004).

Are Randomised Controlled Trials Really the 'Gold Standard' in Restorative Justice Research?

- Strang, H., Barnes, G., Braithwaite J. and Sherman L. W. (1999) *Experiments in Restorative Policing: A Progress Report on the Canberra Reintegrative Shaming Experiments (RISE)*. Available online at: <http://www.aic.gov.au/rjustice> (12 January 2004).
- Strang, H. and Sherman, L. (2003) *Effects of Face-to-Face Restorative Justice on Repeat Offending and Victim Satisfaction: A Systematic Review for the Campbell Collaboration*. (May 16, 2003).
- Tilley, N. (2001) 'Evaluation and Evidence-led Crime Reduction Policy and Practice', in R. Matthews and J. Pitts (eds.) *Crime, Disorder and Community Safety: A New Agenda*. London: Routledge.
- Tyler, T. (1990) *Why People Obey the Law*. New Haven: Yale University Press.
- Vennard, J., Sugg, D. and Hedderman, C. (1997) *Changing Offenders' Attitudes and Behaviour: What Works*. Home Office Research Study 171. London: Home Office.
- Wilcox, A. (2003) 'Evidence-based Youth Justice? Some Valuable Lessons from an Evaluation for the Youth Justice Board', *Youth Justice* 3(1):19-33.
- Wilcox, A. with Hoyle, C. (2002) *Final Report for the Youth Justice Board on the National Evaluation of Restorative Justice Projects*. Unpublished report. London: Youth Justice Board.
- Young, R. and Hoyle, C. (1999) *Restorative Cautioning: Strengthening Communities in the Thames Valley. Confidential Interim study*. Oxford: Centre for Criminological Research, University of Oxford.
- Young, R. and Hoyle, C. (2000) 'Examining the Guts of Restorative Justice', *Criminal Justice Matters* 40: 33-4.
- Young, R. (2001) 'Just Cops doing Shameful Business', in A. Morris and G. Maxwell (eds.) *Restorative Justice for Juveniles*. Oxford: Hart Publishing.