

Improving the evaluation of crime prevention and reduction programs through research-practitioner partnerships

Anthony Morgan

Australian Institute of Criminology

Anthony.morgan@aic.gov.au

Keywords: Researcher-practitioner partnerships, Evaluation stakeholders, Stakeholder involvement, Crime prevention, Crime reduction, Criminal justice, Quasi-experimental design, Scientific methods.

Abstract

While there is widespread recognition of the importance of stakeholder involvement in evaluation, less attention has been given the role of researcher-practitioner partnerships in evaluations using rigorous scientific methods, such as quasi-experimental designs. This is particularly true for the evaluation of crime prevention and reduction programs, including those conducted by evaluators independent of program design and delivery. Reflecting on several evaluations of programs designed to prevent and reduce crime and respond to the needs of vulnerable populations in court settings, this paper highlights the benefits, challenges and lessons from working in partnership with practitioners to conduct rigorous outcome evaluations. While evaluations are often conducted by someone independent of program management and delivery; it is still important for researchers to work in partnership with policy makers, program managers and project staff to ensure evaluations are methodologically rigorous, successfully implemented and focus on delivering practical recommendations for action.

Introduction

The importance of working with stakeholders is now firmly established in evaluation theory and practice (Brandon & Fukunaga, 2014). Involving stakeholders in an evaluation is believed to offer a number of benefits, such as increasing support for evaluation and the use of evaluation findings, and for this reason is a common feature across a range of different approaches to evaluation. Despite this, few areas of evaluation are as divisive as determining the optimal nature and extent of stakeholder involvement in an evaluation, and the level of involvement encouraged therefore varies depending on the approach that is adopted (Hall, 2008).

The aim of this paper is to demonstrate the value of partnerships between researchers and practitioners in those situations in which an external evaluator, independent of program design and delivery, has been commissioned to undertake the evaluation. In particular, this paper argues that the relationship between these external evaluators and agencies commissioning evaluation should be viewed as a partnership, and that working in partnership can increase the likelihood that rigorous scientific methods will be used to measure program effects and that, where they are used, will be more likely to be successfully implemented. The paper begins by reflecting on the standard of crime prevention and criminal justice evaluation in Australia and describing some of the barriers to high quality evaluation. Using several recent evaluations as case studies, the benefits and challenges associated with researcher-practitioner partnerships in evaluations involving quasi-experimental methods are then described. The paper ends by outlining some key lessons for working collaboratively with evaluation stakeholders.

Evaluating strategies to prevent and reduce crime

There has been rapid growth in the evaluation of strategies designed to prevent and reduce crime (National Research Council, 2005; Tilley & Clarke, 2006; Weatherburn, 2005). While there is certainly room for improvement, government agencies are now more likely than ever to allocate funds within program budgets for evaluation research, with a view to commissioning independent evaluators to provide an objective and impartial assessment of the effectiveness, efficiency and appropriateness of policies and programs (New South Wales, Department of Premier & Cabinet, 2013; Western Australia, Program Evaluation Unit, Department of Treasury, 2014). This is reflected in the amount of evaluation activity that is undertaken by government researchers, academics and consultants (as well as internal evaluations conducted by practitioners), the number of researchers working in evaluation, and the level of procurement activity managed by criminal justice authorities.

There are several different approaches to measuring the impact of programs designed to prevent and reduce crime. Selecting an appropriate evaluation design and research method requires consideration of the characteristics of a program, the purpose of the evaluation, the available options, and the views of key stakeholders (English, Cummings & Stratton, 2002; National Research Council, 2005). However, experimental (especially quasi-experimental) and observational methods are the most common approaches used in crime prevention and criminal justice research (Idriss, Jendly, Karn & Mulone, 2010). Given the influence of experimental methods, the Scientific Methods Scale (SMS) was developed to assess the quality of outcome evaluations in crime prevention and criminal justice research. The SMS forms the basis of systematic reviews of crime prevention methods undertaken by the Campbell Collaboration (Farrington, Gottfredson, Sherman & Welsh, 2006; Sherman, Farrington, Welsh & MacKenzie, 2006), while similar criteria have been used by the Washington State Institute of Public Policy (Lee et al. 2012). The SMS is primarily focused on ensuring the highest possible level of internal validity and drawing valid conclusions regarding the causal relationship between interventions and the outcomes observed. The scale ranges from a correlation between a program

and a measure of the outcome (level one) through to randomised control studies (level five), which are widely (but not universally) regarded as the gold standard for evaluation research (Farrington et al. 2006; Tilley & Clarke, 2006). A research design that achieves level three on the SMS, with measures of the outcome (usually a reduction in crime) pre and post intervention and an appropriate comparison group against which to compare results (a quasi-experimental design) is considered the minimum design for drawing valid conclusions about the effectiveness of a strategy (Farrington, et al. 2006; Sherman, Gottsfredson, MacKenzie, Eck, Reuter & Bushway, 1998).

Applying this standard, large-scale systematic reviews have shown that there is an accumulated body of high quality research demonstrating the effectiveness of crime prevention and criminal justice strategies (e.g. Lee et al. 2012; Sherman et al. 1998; Sherman, Farrington, Welsh & MacKenzie, 2006). Most of these reviews have acknowledged the need to improve the standard of evaluation practice, with many studies failing to meet the criteria for inclusion.

However, the number of Australian initiatives included in the systematic reviews, meta-analyses and databases describing effective (and ineffective) interventions are relatively small when compared to other countries (Morgan & Homel, 2013). This is particularly apparent when considered alongside the level of crime prevention activity (as illustrated by the Australian Crime and Violence Prevention Awards and the number of commonwealth, state and territory funding programs), the financial resources invested in the criminal justice system (Steering Committee for the Review of Government Service Provision [SCGRSP], 2014), and the volume of evaluation research that has been conducted. There has only been a handful of randomised control studies conducted in Australia (Lind et al. 2002; Jones, 2011; Mazerolle, Antrobus, Bennett & Tyler, 2013; Sherman, Strang & Woods, 2000). A review of community-based crime prevention strategies suitable for local government to address a number of common crime types in NSW found that, despite the emphasis on the important role of local government in crime prevention in this country for more than two decades, fewer than 20 Australian studies met the criteria for inclusion (Morgan, Boxall, Lindeman & Anderson, 2012). Similarly, a recent research project by the Australian Institute of Criminology exploring the evidence in support of police crime prevention in Australia has also revealed gaps in the amount and quality of evaluation research (Morgan & Mann, forthcoming). There is a large body of Australian evidence for certain responses, including random breath testing (Hendrie, 2003; Shults et al. 2001), community-based and regulatory responses to alcohol-related violence (Morgan, Boxall, Lindeman & Anderson, 2012; National Drug Research Institute, Curtin University of Technology, 2007), cautions and conferencing (Morgan & Mann, forthcoming) and a significant body of research (much of which has been undertaken by the NSW Bureau of Crime Statistics & Research) into the impact of different court initiatives and sentence options.

There are several possible explanations for the variable quantity and quality of Australian studies evaluating the impact of strategies to reduce and prevent crime. It is likely to be a combination of factors such as insufficient funding for robust experimental designs (which can be costly), a reluctance among program managers to subject their programs to rigorous evaluation for fear of negative results, disagreement between evaluators and program managers on questions of attribution, problems accessing meaningful data on

program effects, challenges identifying suitable comparison groups or areas (particularly outside of institutional settings or when evaluating place-based strategies), or the limited skills, knowledge or experience among those entrusted with evaluation (Morgan & Homel, 2013). It may also be due to ideological differences about the best way to assess the effectiveness of strategies to prevent and reduce crime (Tilley & Clarke, 2006).

Improving evaluation standards through partnership approaches

This paper argues that some of these common barriers to rigorous outcome evaluations, particularly evaluation based on quasi-experimental research designs, can be overcome by establishing and maintaining effective partnerships between evaluators and practitioners (in this paper practitioner collectively refers to policy makers, program managers and project staff). This is not a new concept – the importance of working with stakeholders in evaluation is well established (Brandon & Fukunaga, 2014; Plottu & Plottu, 2009). Stakeholder involvement is a common principle underpinning many different approaches to evaluation (Brandon & Fukunaga, 2014). For example, utilisation-focused approaches are underpinned by a commitment to meeting the needs and requirements of the intended user/s (which informs the decision about the evaluation approach and methods) and the participation of these users in decision-making at each stage of the evaluation (Patton 2008). In participatory approaches to evaluation, stakeholders are involved in all aspects of the evaluation and empowered to own the evaluation process, fulfilling roles normally assigned to the evaluator and building evaluation skills and knowledge in the process (Greene, 2006). Reflecting the widespread acceptance of the important role of stakeholders in evaluation, there has even been some acknowledgement of the need to involve stakeholders as part of quantitative approaches to measuring the impact of programs, including areas such as education research (Datta, 2006).

While these approaches differ in terms of the philosophical position they adopt and their rationale for involving stakeholders, there are some common themes in terms of the range of benefits that are believed to result from involving stakeholders in the process of planning an evaluation, collecting and analysing data and reporting findings and recommendations. These include increasing the likelihood that evaluation findings will be used, facilitating access to better quality data and contextual information and enhancing the perceived credibility and validity of the evaluation design, methods and results among key stakeholders (Braga, 2013; Brandon & Fukunaga, 2014; Bryson, Patton & Bowman, 2011; Cullen, Coryn & Rugh, 2011; Patton, 2008; Plottu & Plottu, 2009). Conversely, failing to engage with stakeholders can lead to missed opportunities or result in evaluations that are inaccurate or insensitive to the needs of stakeholders, which may actually deter future investment in evaluation (Bryson, Patton & Bowman, 2011).

Despite the important role of stakeholders in evaluation, and the potential value of researcher-practitioner partnerships, little attention has been given to these arrangements within criminal justice and crime prevention evaluations. For example, a report by the National Research Council (2005) put forward a number of recommendations to improve the evaluation of anticrime programs, with a particular focus on increasing the number of experimental and quasi-experimental designs. While the report argued

that both evaluators and policy makers needed to do more to increase the amount and methodology quality of impact evaluations, little attention was paid on how the two parties might better work together. Similarly, commentators arguing for criminal justice evaluations that are methodologically rigorous have tended to be preoccupied with debates about the relative strength of different research designs and methodological approaches, particularly in terms of maximising internal and external validity (Braga, Welsh & Bruinsma, 2013; Idriss et al. 2010; Tilley & Clarke, 2006). While there have been a number of studies exploring research utilisation by criminal justice agencies (Lum, Telep, Koper & Grieco, 2012), and others exploring the impact of researcher-practitioner partnerships on evaluation findings (Petrosino & Soydan, 2005; Welsh, Braga & Hollis-Peel, 2012), there has been limited research into researcher-practitioner partnerships in criminal justice research and, as a result, little is known about the factors that contribute to successful partnerships for evaluation (Alpert, Rojek & Hansen, 2013). This paper attempts to address this gap.

Benefits of working in partnership with policy makers, program managers and project staff

There are different arrangements in terms of who is responsible for an evaluation and their relationship to the program being evaluated. Petrosino and Soydan (2005) reviewed more than 300 individually-focused crime reduction programs and identified three categories of evaluation teams, each with a number of sub-categories: internal (program developer/creator, program/agency staff, government evaluator), external (academic researchers, private research firm, foundation/not for profit) and collaborative approaches (academic/practitioner, academic/government). The Australian Institute of Criminology (AIC), which is frequently engaged by both Commonwealth and state and territory government agencies to evaluate programs designed to prevent and reduce crime, has typically been engaged as an external evaluator, independent of program design and delivery. As an external evaluator, the AIC is tasked with making an objective and impartial assessment of the impact of the program being evaluated, drawing upon its specialist expertise to design and implement rigorous evaluation methodologies to produce valid, transparent and defensible findings.

In performing this role, researchers involved in these evaluations have frequently approached the relationship with the agency commissioning the evaluation as a partnership. This meant working closely with that agency (and other partners) during the planning stages of the evaluation, data collection and analysis and during the reporting of evaluation findings. Several examples of both published and unpublished evaluations conducted by the AIC are used to describe these partnership arrangements, as well as the benefits that resulted:

- Evaluation of the Queensland Murri Court (Morgan & Louis 2010): An Indigenous sentencing court for both adult and youth defendants that allowed greater input from the Aboriginal and Torres Strait Islander (ATSI) community into the sentencing process.

- Evaluation of the Queensland Special Circumstances Court Diversion Program (SCCDP) (unpublished): A court diversion program for defendants who were homeless or who had a mental illness or intellectual disability.
- Evaluation of alternative dispute resolution initiatives in the care and protection jurisdiction of the NSW Children's Court (Morgan, Boxall, Terer & Harris, 2012): A new model of dispute resolution involving the use of conciliation and mediation to resolve child protection disputes before the Children's Court.
- Evaluation of Indigenous drug and alcohol treatment programs (unpublished): Six community-based residential rehabilitation programs in four jurisdictions designed to reduce substance misuse, particularly among ATSI people.
- Policing licensed premises in the ACT (Smith, Morgan & McAtamney, 2011): A strategy to reduce alcohol-related crime in the ACT's main entertainment precinct, involving intelligence gathering, education for licensees, proactive enforcement of liquor licensing and high visibility policing.
- Partnership between the AIC and CrimTrac (recently commenced): A new program of research focused on the evaluation of information systems and services for law enforcement agencies.

In the majority of these evaluations, the AIC was commissioned through a procurement process to conduct an evaluation of the program, which had important implications for how the evaluations were managed and the nature of partnership arrangements. For this reason, most of the following discussion centres on partnerships between practitioners and evaluators independent of program design and delivery.

Planning the evaluation

Identifying and involving evaluation stakeholders early in the process of planning an evaluation provides a range of benefits, irrespective of the approach to evaluation being used (Brandon & Fukunaga, 2014). However, there are certain benefits that are particularly relevant to the types of evaluations that are the focus of this paper. First, stakeholders can assist with determining what outcomes can and should be attributed to the program being evaluated and, therefore, which outcomes should be the focus of an evaluation. This helps guide the evaluation, determine appropriate evaluation questions, and ensures that there is agreement regarding the scope of the evaluation well in advance of data collection and reporting (avoiding future disagreements). Involving stakeholders in discussions about how a particular program and its component parts contribute to desired outcomes can also empower and encourage the meaningful participation of stakeholders in evaluation, which can 'dramatically increase the chances that the evaluation will meet utility, feasibility, propriety and accuracy evaluation standards' (Donaldson & Lipsey, 2006, pp. 65-66).

One way of approaching this is to involve stakeholders in the development of a program logic model and evaluation framework. A program logic model describes the main activities that will be delivered as part of a program, and the relationship between these activities and the hierarchy of short, intermediate and long-term outcomes. This helps to determine what outcomes can be reasonably attributed to the program and makes explicit the underlying theory about how a program contributes to these outcomes (Funnell & Rogers, 2011) – hypotheses that can be tested using rigorous scientific methods. In addition to being a powerful communication tool for criminal justice evaluation (Willis & Tomison, 2014, this issue), the logic model also forms the basis of an evaluation framework, which details the specific evaluation questions and performance indicators that will be measured (and how) as part of the evaluation. This enables stakeholders to have input into the design of the evaluation, including data collection processes, which can encourage buy in, facilitate access to data and identify and address potential barriers to conducting the evaluation.

This approach has been used extensively in AIC evaluations. Undertaken in consultation with key stakeholders, the development of program logic models has helped to resolve important conceptual questions (and concerns) about attribution. In an evaluation of six Indigenous drug and alcohol treatment programs, the logic model that underpinned the evaluation (developed with input from program operators) was used to demonstrate the hypothesised link between improved health and wellbeing – the direct consequence of effective treatment – and reduced recidivism. This was particularly helpful in convincing key stakeholders of the need to measure reoffending for programs that did identify this as an explicit objective. Using a quasi-experimental research design (comparing program completers and non-completers), the evaluation concluded that there was evidence that treatment had a positive impact on reoffending, helping to address a significant gap in the available evidence base. More recently, the use of a program logic approach has been used to determine what outcomes could be attributed to a number of law enforcement data services provided by CrimTrac, and persuade policing agencies of the need to measure these important outcomes. These logic models were developed with considerable input from both CrimTrac and their police agency partners.

Involving stakeholders in the evaluation planning process can also help to overcome resistance to the use of quasi-experimental research designs and the challenges associated with identifying suitable comparison areas. In the evaluation of alternative dispute resolution for care and protection matters in the NSW Children's Court, the evaluators worked closely with the evaluation working group to develop an innovative solution to challenges associated with determining an appropriate comparison group comprising matters that did not go through the alternative model of decision making. This was necessary because the program was rolled out state wide (i.e., there was no natural comparison area). Given the significance of the reforms, which were introduced in response to a Royal Commission, the working group was reluctant to restrict access to the program group for the purpose of evaluation. The research team therefore negotiated a retrospective comparison group, including appropriate parameters and a process for collecting additional data for this purpose. The final report on the evaluation compared the outcomes for matters referred to ADR and this comparison group, including a cost savings comparison, and the findings from this analysis were instrumental in informing a recommendation to continue to support the use of ADR – a recommendation that was

supported by the then Attorney General (Morgan, Boxall, Terer & Harris, 2012).

Similarly, in the evaluations of the Queensland Murri Court (Morgan & Louis, 2010) and the Queensland Special Circumstances Court Diversion Program, the evaluators worked closely with the Department that commissioned the evaluation and the program Steering Committee (comprising representatives from each agency involved in the program) to identify a suitable comparison group of offenders who did not participate in the program. There was resistance to this approach among some stakeholders – due largely to a limited understanding of the rationale for the comparison group – and various practical barriers to accessing data on offenders in the comparison group, but these were overcome through careful negotiation, by explaining both the strengths and limitations of the different options, and working collaboratively with court partners to develop an appropriate solution. Working with these stakeholders during these early stages to determine the comparison group was important because these same stakeholders were responsible for reviewing the findings from the evaluation. Had the evaluators simply decided on the eligibility criteria for the comparison group in isolation and without consultation, the process of collecting the data and the results from any analysis would likely have been met with significant resistance and scepticism. As it was, the collaborative and consultative approach was particularly helpful given the eventual findings from the evaluation of the Queensland Murri Court (which is discussed later).

Conducting the evaluation

A collaborative approach to evaluation recognises the potential of involving program managers and staff in the data collection process, rather than only seeing them as sources of data for the purpose of the evaluation (Cullen, Coryn & Rugh, 2011). Involving program staff in this way can help to increase the perceived credibility and validity of the findings. It is common for there to be significant gaps in the data required to measure the outcomes from crime prevention and criminal justice programs. Administrative databases – where they exist – are used extensively to measure the impact of criminal justice programs. However, these databases are often not established for the purpose of evaluation, and there are limitations to their use. For example, the evaluations of the Queensland Murri Court and the Queensland Special Circumstances Court Diversion Program (SCCDP) were undertaken as part of long-term partnership with the Department of Justice and Attorney General, and followed earlier evaluations of the Queensland Drug Court. Recognising the need to utilise court data to measure the impact of the Queensland Murri Court, and to integrate court data with other information on important risk factors for future offending, the Department developed a new system for recording information on all court participants and the comparison group. The AIC was instrumental in helping to design this database, working closely with the Department to build a system that would eventually be used for all court innovation programs across the state. This was a time consuming process, but was critical in ensuring that adequate data were available to measure the impact of the Murri Court and Queensland SCCDP (and other programs) on recidivism.

In the evaluation of alternative dispute resolution (ADR) in the NSW Children's Court, there was no pre-existing administrative database for care and protection matters. This meant that initially data were not available to measure the impact of ADR in the time required to finalise care matters or the proportion of matters that resulted in a court hearing. The AIC worked very closely with the evaluation working group to establish a process for extracting information from the hardcopy case files in a systematic way, and to negotiate a process by which the information was extracted by program staff. This could only have been achieved by working in partnership with the evaluation working group, and ensured access to data for the analysis of key outcomes and cost-savings.

Further, as part of this same evaluation, a number of other stakeholders were involved in data collection. Conference chairs (registrars and mediators) completed a report on the outcomes of each conference for the purpose of the evaluation, and also disseminated a brief satisfaction survey to conference participants – whom the evaluators would have struggled to survey and achieve a similar response rate. In the evaluation of policing responses to alcohol related violence in and around licensed premises in the Australian Capital Territory (ACT), the AIC worked with ACT Policing to develop and implement a 'place of last drink' form for all alcohol-related incidents attended by police. In the evaluation of the Queensland SCCDP and evaluation of Indigenous drug and alcohol treatment programs, case managers completed a client questionnaire designed by the AIC at program entry and exit to measure change in the health and wellbeing of participants. In each of these examples, program staff had better access to the information source (i.e. program participants) than the evaluators and were able to capture information that would have far more difficult to collect otherwise. Where possible, these evaluation processes were embedded as part of the design of the program, thereby minimising the additional impost on program staff. This also requires that appropriate data collection protocols be developed and agreed to ensure the validity and reliability of the data being collected by third parties.

Reporting on findings

Working in partnership with relevant stakeholders as part of the process of reporting findings from the evaluation can help to overcome a major limitation of experimental and quasi-experimental research designs – namely, that they provide evidence of whether an outcome has been achieved, but not the reasons why the outcome has or has not been observed. By providing stakeholders with an opportunity to consider and respond to preliminary findings, such as by holding a workshop with these stakeholders prior to submitting the final report, it is possible to identify alternative explanations or additional information that may be crucial to an understanding of the results. This is particularly useful where the results are unexpected or potentially negative. In the example of the Queensland Murri Court, initial results showed a negative finding – offenders sentenced in Murri Court were more likely to be sent to prison. Following extensive consultation with the evaluation working group, and working with the Department to re-examine the data, a valid explanation was identified. Offenders appearing before the Murri Court were more likely to be serving a term of imprisonment for earlier offences at the time of being sentenced, which meant they were also more likely to receive a further custodial

penalty (i.e., an increase to the original imprisonment term). Once the data were re-analysed using modified parameters to take this issue into account, the true impact of the Murri Court on sentencing outcomes could be determined (although the final result was still not entirely positive as offenders sentenced in the Murri Court were still just as likely as the comparison group to receive a custodial sentence). This approach minimised any delay and ensured a resolution could be identified quickly, before the results were disseminated to a broader audience. Importantly, experience shows that evaluation stakeholders are less likely to question or criticise the technical quality of an evaluation where they have been involved in the process, even if the findings are unexpected or negative (Brandon & Fukunaga, 2014). This means any discussion around results can focus on explaining and understanding the findings – as was the case with Murri Court. In the end, the results of the evaluation were mixed, and input from a range of stakeholders, particularly representatives from the Indigenous community, was used to help explain the findings, as well as to measure a number of important outcomes that could not easily be quantified, such as the impact on the partnership between the court and Indigenous community.

Another opportunity for collaboration at the time of reporting and communicating findings is in developing evaluation recommendations. While evaluators must control the process, ensure that any recommendations are substantiated by evidence from the evaluation and give adequate consideration to the likely reaction to the recommendations from different stakeholders (Bryson, Patton & Bowman, 2011), working with program staff to develop recommendations can increase the likelihood that the findings from an outcome evaluation will be used to inform decision making. In the evaluation of ADR in the NSW Children's Court, a workshop was held with members of the evaluation working group to discuss and agree on the proposed recommendations (which were provided in advance of the workshop). Much of this discussion was focused on the reasons that one of the two models evaluated was not as effective as the other (based on results from the quasi-experimental evaluation) and potential strategies for how that model might be improved. Rather than attempt to influence these recommendations in a negative way, this approach resulted in recommendations that were more practical, useful and offered a clearer course of action.

Limitations associated with research-practitioner partnerships for evaluation

The discussion above described a range of benefits associated with adopting a partnership approach to evaluating strategies to prevent and reduce crime using rigorous scientific methods. However, there are limitations and challenges associated with this approach that need to be acknowledged. First, collaborative approaches to evaluation are time consuming and resource intensive, requiring additional resources to manage the partnership and a commitment from program staff to dedicate additional resources to support evaluation processes (Brandon & Fukunaga, 2014). In the examples described throughout this paper, significant time was devoted to establishing effective partnership arrangements with the representatives of the agency commissioning the evaluation (and other partners, where necessary) – although there was a substantial return on this investment in terms of ensuring the successful implementation of the evaluation.

The most obvious and perhaps important criticism of researcher-practitioner partnerships is the actual or perceived impact on the objectivity and validity of the research findings. Views regarding the appropriateness of stakeholder involvement in evaluation employing experimental and quasi-experimental research designs are mixed (Brandon & Fukunaga, 2014; Pollitt, 2009). Some authors argue that the two are not mutually exclusive and that evaluators should involve stakeholders in the evaluation process (Patton, 2008; Plottu & Plottu, 2009), while others have suggested that there are occasions in which evaluations based on rigorous scientific methods should be conducted independently of program staff (Pollitt, 2009). There is an inherent tension that comes with being commissioned by the people responsible for the design, management and/or implementation of a program to conduct the evaluation. Providing an objective and independent assessment of the success (or not) of that program can be a source of conflict between program managers and evaluators, and pressure to adjust or soften unfavourable findings is not uncommon (Tilley & Clarke, 2006). While previous research into participatory approaches to evaluation has suggested they enhance the validity and credibility of the findings (at least among the stakeholders involved in the evaluation), they also acknowledge the potential for bias (Brandon & Fukunaga, 2014; Cullen, Coryn & Rugh, 2011).

The findings from empirical studies on the impact of researcher-practitioner partnerships on the technical quality and potential for bias in evaluations of strategies designed to prevent and reduce crime have been mixed (Petrosino & Soydan, 2005; Welsh, Braga & Hollis Peel, 2012). Petrosino & Soydan (2005) reviewed 12 meta-analyses on offender treatment programs and found that all but one of these previous studies had observed higher effect sizes when evaluators were involved in the design or delivery of the program. They also conducted their own meta-analysis of almost 300 randomised control trials of individually focused crime reduction and observed a similar result. Importantly, effect sizes were also higher for evaluations that involved collaboration between researchers and practitioners (i.e., where evaluators were not directly involved in program delivery). However, Welsh, Braga and Hollis-Peel (2012) conducted a meta-analysis of more than 40 experimental and quasi-experimental studies into the effects of police crime prevention strategies and did not find a relationship between evaluator involvement and program effects. Two explanations have been provided for the possible relationship between evaluator involvement and program effects – the high fidelity theory, which argues that involving researchers in program design and delivery may increase program fidelity, and the cynical view, which suggests there is pressure on the evaluator to deliver positive results and therefore some level of bias (intentional or unintentional) (Petrosino & Soydan 2005; Welsh, Braga & Hollis-Peel, 2012). While the cynical view cannot be discounted (Eisner, 2009), and past research has been less concerned with partnership arrangements than with evaluator involvement in program design and delivery, it is still important that steps be taken to minimise the risk of bias. This might involve strategies such as offering the opportunity for external oversight of an evaluation by an independent researcher (Braga, 2013), and ensuring the evaluator remains in control of the evaluation process (Cullen, Coryn & Rugh, 2011).

A balanced approach to evaluation

The focus of this paper has been on improving the standard of outcome evaluations for strategies designed to prevent and reduce crime through researcher-practitioner partnerships. The emphasis has been on increasing the use of rigorous scientific methods to measure program effects. However, adopting a quasi-experimental design to evaluate programs does not preclude the use of qualitative data collection methods, which are another mechanism for engaging stakeholders as part of the evaluation. In all of the evaluations described in this paper, a mixed methods approach was adopted, combining the quantitative measurement of key outcomes (e.g. reoffending) with qualitative methods. Combining quasi-experimental research designs with field study (e.g. interviews) is a common approach to understanding social change, particularly among pragmatists (see Boxall, this issue), and offers a number of important benefits (Hall, 2008). First, these interviews can confirm (through triangulation) the findings from the quantitative analysis, which can add weight and credibility to the findings. Second, they can help to provide explanation and context for the findings from any quantitative analysis of key outcomes. Third, these interviews can address evaluation questions that cannot be answered using quantitative methods. Fourth, these interviews will usually be a primary source of information required for a process evaluation and developing a comprehensive understanding of the program being evaluated, which will directly inform recommendations. Fifth, incorporating these qualitative approaches can help to achieve buy in from stakeholders who are less supportive of quantitative methods. And finally, qualitative interviews can help to address the common criticisms of quasi-experimental methods for lacking external validity by developing a more detailed understanding of the mechanisms that underpin interventions and the context in which these mechanisms are applied (Tilley & Clarke, 2006).

Factors contributing to successful partnership arrangements for evaluating strategies to prevent and reduce crime

Partnership approaches to evaluation are not that dissimilar to partnerships in other contexts, including those that exist as part of program delivery. A recent review of empirical studies exploring stakeholder involvement in evaluation identified many of the criteria for effective partnership working as being relevant to evaluation (Brandon & Fukunaga, 2014). Therefore in thinking about how best to manage the relationship between evaluators and the agencies that commission an evaluation, it is helpful to reflect on the qualities of effective partnerships and governance arrangements required (to varying degrees, depending on the nature of the partnership) for partnerships to function effectively.

Table 1: **Criteria for effective partnerships and their relevance to evaluation**

<i>Criteria for effective partnership working</i>	<i>Relevance to evaluation</i>
A clear mission and agreement on the objectives of the partnership	Evaluations were less likely to encounter resistance or practical barriers to accessing data when evaluators and program staff shared similar goals for the evaluation, particularly in terms of making a valid assessment of the effectiveness of the program.
Good knowledge and understanding of one another's roles and responsibilities	Agreement on roles and responsibilities for all parties involved in the evaluation at the commencement of the evaluation, and documenting this in either a project implementation plan or contract, helped to ensure that there was no confusion about who was responsible for undertaking evaluation activities, particularly with regards to data collection.
A high level of trust between partner agencies, including members that work well together, respect one another and are committed to ensuring the partnership succeeds	The examples in this paper involved evaluators and program staff that had established a positive relationship built on trust. This helped to minimise any suspicion among program staff about the motives of evaluators pressing for more rigorous designs and overcome reluctance to share data and support the evaluation.
Strong leadership, including local 'champions'	In the evaluation of the Murri Court and SCCDP and the evaluation of ADR in the NSW Children's Court, there was a champion in the agency that had commissioned the evaluation who advocated for the evaluation and encouraged other stakeholders to participate. This was a key factor in their success.
Adequate resourcing, including staff having enough time away from agency core business to provide input to the partnership	In each of the examples presented in this paper, evaluators were adequately resourced to undertake a high quality evaluation. However, for collaborative approaches to evaluation to function effectively, program staff also needed to be able to invest the necessary time to undertake tasks to support the evaluation, including data collection.

Table 1: **Criteria for effective partnerships and their relevance to evaluation** *Continued*

<i>Criteria for effective partnership working</i>	<i>Relevance to evaluation</i>
Partnership structures that are relatively small, businesslike, with a clear process for making decisions and a focus on problem solving	As described throughout this paper, a working group comprising representatives from the various parties was established early in each evaluation to oversee the development, implementation and ongoing review of the evaluation. This group was responsible for facilitating access to data and personnel, providing input on key outputs as they were produced, providing input into the final report recommendations and developing solutions to any problems (such as access to data), as they arose.
Data sharing policies and protocols	The evaluations described in this paper required access to data from a range of sources, and it was important to ensure the privacy and confidentiality of that data. There needed to be clear policies and protocols for the sharing of information between the evaluators and program staff (typically managed as part of a contract). Where program staff were collecting data on behalf of the evaluators, there were processes in place to ensure the data collected were both valid and reliable.
Continuity in partner representation and participation and documentation of processes and decision-making	Program turnover can affect the progress of evaluations as well as program delivery. While not ideal, this was encountered in a number of evaluations. Where there was appropriate documentation and effective transition arrangements, the impact on the evaluation was minimal.

Source: Gilling, 2005; Homel, 2006; Morgan & Homel, 2011; Rosenbaum, 2002.

Table 1 describes a number of criteria for effective partnership working in crime prevention and criminal justice settings, based on an established body of literature. The relevance of these criteria to evaluations employing rigorous scientific methods is also highlighted, based on the examples discussed in this paper. This shows that for researcher-practitioner partnerships to work effectively as part of an evaluation, there needs to be clear agreement regarding the use of quasi-experimental research designs (or rigorous scientific methods more broadly), clear roles for the various parties in supporting the evaluation, a high level of trust between the evaluator and agency commissioning the evaluation and adequate resourcing to support the approach to evaluation.

Conclusion

This paper began by arguing that researcher-practitioner partnerships can help improve the design and implementation of evaluations conducted by evaluators independent of program design and delivery using rigorous scientific methods. Then, reflecting on several evaluations of programs designed to prevent crime and respond to the needs of vulnerable populations in court settings, the paper identified a number of important benefits that result from working in partnership with policy makers, program managers and project staff. Involving stakeholders can assist with:

- determining what outcomes can and should be attributed to the program being evaluated and should therefore be the focus of an evaluation;
- designing the evaluation, including data collection processes, which can encourage buy in, facilitate access to data and identify and address potential barriers to conducting the evaluation;
- overcoming resistance to the use of quasi-experimental research designs and the challenges associated with identifying suitable comparison areas;
- providing assistance with the collection of data, particularly from hard to reach populations, to enable key outcomes to be measured;
- identifying alternative explanations or additional information that may be crucial to an understanding of the results; and
- developing recommendations that are more practical, more useful and offer a clear course of action to make improvements to the program.

There are certain challenges and limitations associated with this approach, but experience has shown that these are far outweighed by the benefits described above and that problems such as bias may not be as common as sometimes believed.

In an era of growing emphasis on accountability, government agencies will continue to engage evaluators independent of program design and delivery to make an objective and impartial assessment of the impact of crime prevention and criminal justice programs. Experimental research designs (particularly quasi-experimental designs) are also likely to remain the sought after standard for many evaluations of strategies to prevent and reduce crime. Evaluators involved in evaluations that employ quasi-experimental designs would benefit greatly from embracing the importance of working cooperatively with commissioning agencies, program staff and other key stakeholders, and viewing the relationship as a partnership. There is, after all, very good reason that stakeholder involvement and participation has become such a dominant theme in evaluation theory and practice.

References

- Alpert, G. P., Rojek, J., & Hansen, J.A. (2013). *Building bridges between police researchers and practitioners: Agents of change in a complex world*. Washington DC: US Department of Justice.
- Brandon, P. R. & Fukunaga, L. L. (2014). The state of the empirical research literature on stakeholder involvement in program evaluation. *American Journal of Evaluation*, 35(1), 26-44.
- Braga, A.A. (2013). Embedded criminologists in police departments. *Ideas in American Policing*, No. 17, pp. 1-20.
- Braga, A. A., Welsh, B. C. & Bruinsma, G. J. N. (2013). Integrating experimental and observational methods to improve criminology and criminal justice policy. In B. C. Welsh, A. A. Braga & G. J. N. Bruinsma (Eds.). *Experimental criminology: Prospects for advancing science and public policy* (pp. 277-298).
- Bryson, J. M., Patton, M. Q., & Bowman, R. A. (2011). Working with evaluation stakeholders: A rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, 34, 1-12.
- Cullen, A. E., Coryn, C. L. S., & Rugh, J. (2011). The politics and consequences of including stakeholders in international development evaluation, *American Journal of Evaluation*, 32(3), pp. 345-461.
- Datta, L. (2006). The practice of evaluation: Challenges and new directions. In I. F. Shaw, J. C. Greene & M. M. Mark (Eds.). *The Sage Handbook of Evaluation*. (pp.419-438). London: Sage Publications.
- Donaldson, S. L., & Lipsey, M. W. (2006). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In I. F. Shaw, J. C. Greene & M. M. Mark (Eds.). *The Sage Handbook of Evaluation*, (pp. 56-75). London: Sage Publications.
- Eisner, Manuel. (2009).. No effects in independent prevention trials: Can we reject the cynical view? *Journal of Experimental Criminology*, 5 (2), pp. 163–184.
- English, B., Cummings, R., & Stratton, R. (2002). Choosing an evaluation model for community crime prevention programs. In N. Tilley (Ed.). *Evaluation for crime prevention* (pp. 119-169). Monsey, NY: Criminal Justice Press.
- Farrington, D. P., Gottfredson, D. C., Sherman, L.W., & Welsh, B. C. (2006). The Maryland Scientific Methods Scale. In L. W. Sherman, D. P. Farrington, B. C. Welsh & D. L. MacKenzie (Eds.). *Evidence-based crime prevention* (pp. 13-21). London: Routledge.
- Funnell, S., & Rogers, P. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco: Jossey Bass.

- Gilling, D. (2005). Partnerships and crime prevention. In N. Tilley (Ed.). *Handbook of crime prevention and community safety* (pp. 734-756). Cullompton, UK: Willan Publishing.
- Greene, J. C. (2006). Evaluation, democracy and social change. In I. F. Shaw, J. C. Greene & M. M. Mark (Eds.). *The Sage Handbook of Evaluation* (pp. 118-140). London: Sage Publications.
- Hall, R. (2008). *Applied social research: Planning, designing and conducting real-world research*. South Yarra: Palgrave MacMillan
- Hendrie, D. (2003). *Random breath testing: Its effectiveness and possible characteristics of a 'best practice' approach*. Crawley: The University of Western Australia.
- Hommel P. (2006). Joining up the pieces: what central agencies need to do to support effective local crime prevention. In J. Knutsson & R. Clarke (Eds.), *Putting theory to work: Implementing situational prevention and problem-oriented policing* (pp. 111-139). New Jersey: Prentice Hall.
- Idriss, M., Jendly, M., Karn, J., & Mulone, M. (2010). *International report on crime prevention and community safety: Trends and perspectives, 2010*. Montreal: International Centre for the Prevention of Crime.
- Jones, C. (2011). Intensive judicial supervision and drug court outcomes: Interim findings from a randomised control trial. *Crime and Justice Bulletin*, No. 152, pp. 1-16.
- Lee, S., Aos, S., Drake, E., Pennucci, A., Klima, T., Miller, M., Anderson, L., Mayfield, J., & Burley, M. (2012). *Return on investment: Evidence-based options to improve statewide outcomes*, April 2012, Olympia, Washington: Washington State Institute for Public Policy.
- Lind, B., Weatherburn, D., Chen, S., Shanahan, M., Lancasar, E., Haas, M., & De Abreu Lourenco, R. (2002). *New South Wales Drug Court Evaluation: Cost-effectiveness*. Sydney: NSW Bureau of Crime Statistics and Research, Attorney General's Department. Retrieved from NSW Bureau of Crime Statistics and Research, Attorney General's Department website: <http://www.bocsar.nsw.gov.au/agdbasev7wr/bocsar/documents/pdf/l15.pdf>
- Lum, C., Telep, C. W., Koper, C. S., & Grieco, J. (2012). Receptivity to research in policing. *Justice Research and Policy*, 14(1), pp. 61-95.
- Mazerolle, L., Antrobus, E., Bennett, S., & Tyler, T. R. (2013). Shaping citizen perceptions of police legitimacy: A randomised field trial of procedural justice, *Criminology*, 51(5), pp. 33-63.

- Morgan, A., Boxall, H., Lindeman, K., & Anderson, J. (2012). *Effective crime prevention strategies for implementation by local government*. Canberra: Australian Institute of Criminology (AIC).
- Morgan, A., Boxall, H., Terer, K., & Harris, N. (2012). *Evaluation of alternative dispute resolution initiatives in the care and protection jurisdiction of the NSW Children's Court*. Canberra: Australian Institute of Criminology (AIC).
- Morgan, A., & Homel, P. (2013, July). Evaluating crime prevention: Lessons from large-scale crime prevention programs. *Trends and Issues in Crime and Criminal Justice Series No. 458*, pp. 1-12. Canberra: Australian Institute of Criminology, Australian Government.
- Morgan, A., & Homel, P. (2011). *Model performance framework for local crime prevention*, Canberra: Australian Institute of Criminology (AIC).
- Morgan, A., & Louis, E. (2010). *Evaluation of the Queensland Murri Court: Final report*, Canberra: Australian Institute of Criminology (AIC).
- Morgan, A., & Mann, M. (in press). Police crime prevention in Australia: Findings from a review. *Trends and Issues in Crime and Criminal Justice*.
- National Drug Research Institute, Curtin University of Technology. (2007). *Restrictions on the Sale and Supply of Alcohol: Evidence and Outcomes*, Perth, WA: Curtin University of Technology.
- National Research Council. (2005). *Improving Evaluation of Anticrime Programs*. Washington, DC: The National Academies Press. Available from http://www.nap.edu/catalog.php?record_id=11337
- New South Wales (NSW), Department of Premier & Cabinet. (2013). *NSW Government evaluation framework*. Sydney: NSW Department of Premier and Cabinet.
- Patton, M. Q. (2008). *Utilisation-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology*, 1(4), pp. 435-450.
- Plottu, B., & Plottu, E. (2009). Approaches to participation in evaluation: Some conditions for implementation. *Evaluation*, 15(3), pp. 343-359.
- Pollitt, C. (2009). Stunted by stakeholders? Limits to collaborative evaluation. *Public Policy and Administration*, 14(2), pp. 77-90.
- Western Australia, Program Evaluation Unit, Department of Treasury. (2014). *Evaluation Guide*, Department of Treasury, Government of Western Australia, Perth: Author.

- Rosenbaum, D. (2002). Evaluating multi-agency anti-crime partnerships: Theory, design and measurement issues. In N. Tilley & N. J. Monsey (Eds.). *Evaluation for Crime Prevention*. Criminal Justice Press/Willow Tree Press. Retrieved from National Criminal Justice Reference Service Library website <https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=195631>
- Sherman, L. W., Farrington, D. P., Welsh, B. C., & MacKenzie, D. L. (2006). *Evidence-based crime prevention* (2nd ed.), London: Routledge.
- Sherman, L.W., Gottfredson, D., MacKenzie, D.L., Eck, J., Reuter, P., & Bushway, S. (1998). *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, DC: US Department of Justice.
- Sherman, L. W., Strang, H., & Woods, D. J. (2000). *Recidivism patterns in the Canberra Reintegrative Shaming Experiments (RISE)*. Canberra: Australian Institute of Criminology.
- Shults, R. A., Elder, R. W., Sleet, D. A., Nichols, J. L., Alao, M. O., Carande-Kulis, V. G., ... Task Force on Community Preventive Services. (2001). Reviews of evidence regarding interventions to reduce alcohol-impaired driving. *American Journal of Preventive Medicine*, 21(4), 66-88.
- Smith, L., Morgan, A., & McAtamney, A. (2011). *Policing licensed premises in the Australian Capital Territory*. Canberra: Australian Institute of Criminology (AIC).
- Steering Committee for the Review of Government Service Provision [SCRGSP], (2014). *Report on Government Services 2014*, Canberra: Productivity Commission, Commonwealth of Australia.
- Tilley, N., & Clarke, A. (2006). Evaluation in criminal justice. In I. F. Shaw, J. C. Green & M. M. Mark (Eds.), *The Sage Handbook of Evaluation*. (pp.512-535). London: Sage Publications.
- United Nations Evaluation Group [UNEG], (2005). *Norms for evaluation in the UN System*, Vienna: UNEG, Author.
- Weatherburn, D. (2005). Critical criminology and its discontents: A response to Travers, critique of criminal justice evaluation. *Australian and New Zealand Journal of Criminology*, 38(3), pp. 416-420.
- Welsh, B. C., Braga, A. A., & Peel, M. E. (2012). Can 'disciplined passion' overcome the cynical view? An empirical inquiry of evaluator influence on police crime prevention program outcomes. *Journal of Experimental Criminology*, 8 (4), pp. 415-431.

