
INTRODUCTION: EVALUATION FOR CRIME PREVENTION

by

Nick Tilley

Nottingham Trent University

Evaluation in crime prevention is almost universally called for. It is widely required by funding bodies, and, as a result, it is almost as widely attempted. It is technically very tricky both in relation to specific initiatives (Ekblom, 1990) and in relation to the sorts of partnership approach that are now widely promoted (Rosenbaum, this volume). Critics are agreed that standards are generally very low (Sherman et al., 1997; Ekblom and Pease, 1995; HMIC, 2000; Scott, 2001). There are some fundamental debates about how evaluation should be construed and conducted (Pawson and Tilley, 1997). There are also diverse stakeholders in evaluation, and a range of uses to which evaluations can be put.

The papers in this volume are concerned primarily with the design and conduct of evaluations that are intended to help improve crime prevention policy and practice. Other possible purposes — for example to justify programs, to account for public expenditure, to celebrate achievement, or to encourage greater participation in crime prevention — are not much at issue here.

Much literature on crime prevention has stressed the importance of evaluation for program improvement. The preventive process described by Ekblom (1988), the action research approach advocated by Clarke (1997), and the SARA (scanning, analysis, response, assessment) steps developed by Eck and Spelman (1987) to help operationalise Goldstein's (1979) problem-oriented policing all include evaluation as a key element. The purpose of evaluation in all cases is to provide feedback that will generate corrections to and refinements in crime prevention theory, policy and practice. The University of Maryland report to the United States Congress on evaluations in crime prevention (Sherman et al., 1997) and the British Home Office

report on findings of evaluation studies (Goldblatt and Lewis, 1998) were both designed to inform improvements in what was done in an effort to reduce crime. Likewise, the emerging international Campbell Collaboration has a strong criminal justice stream, which promises systematic reviews of crime prevention with the aim of improving the choice of interventions (Farrington and Petrosino, 2001).

The bottom line in evaluation for crime prevention, is of course, crucial. Evaluation results should allow us more effectively and efficiently to lessen crime and its effects, and we need to measure outcomes in these terms. A number of tricky technical challenges confront the measurement of outcome effectiveness. These include, for example, discounting regression-to-the-mean effects, establishing a secure counterfactual, attributing responsibility for observed changes to the measures introduced, identifying active ingredients in packages of measures, identifying short and long-term effects, identifying significant side-effects, and attaching figures to costs and benefits of interventions. Though these technical challenges need to be met adequately if evaluations are to be useful, that is not sufficient. Evaluations need, preferably at the same time, to produce results that can be generalised.

Science, Theory and External Validity

If evaluation findings are to be useful beyond the program to which they relate, they need to enjoy external validity, and traditional "gold standard" methods may not best deliver this. History, crime investigation and auditing are all activities that require rigour but do not require external validity. In history we need to know how particular events of the past unfolded. In crime investigation, we need to know who committed a particular crime or series of crimes. In auditing we need to know what happened to the turnover, profits and losses of a particular enterprise over a particular period. Both history and crime investigation may from time to time use science, where it is helpful. Auditing aspires to be rigorous in its standards and techniques, but has no pretensions to being a science. Science is concerned with making general statements about regularities and about where and how they are produced. It develops theories and rigorously tests hypotheses better to understand how the natural or social world works. Where theories have been sufficiently tested then they can be applied.

To be most useful — to have external validity — evaluations need to go beyond mere technical rigour. They need to be scientific. This means they need to be driven by theories about how and where changes are produced. Moreover these theories need to be tested.

Decisions about the extension, continuation, alteration or cancellation of a program all depend on findings being applicable beyond the particulars of past events in particular places.

There are some corollaries, of adopting a scientific approach, which will be very challenging. Science deals in unobservable causal powers (for example natural selection, gravity), that may challenge common sense (Bhaskar, 1974). It does not always use the categories available to common sense. Instead it looks to regularities and causal processes that lie beneath the surface, and on this basis reconstitutes how we classify. Just as post-Darwinian classifications of animals and plants go beyond mere resemblance to look to real distinctions by origins, so too the classification of crime prevention measures will change with a better grasp of how they work. Clarke's classification of crime prevention measures by the ways in which they are deemed to operate is a case in point (Clarke, 1997). It is rooted in a theory of how patterns of crime generation are altered. The policy and practitioner communities, working alongside evaluators, need to learn not to ask crass questions, for example, about whether or not, for example, Neighbourhood Watch or Block Watch "works." The diverse activities that may or may not be involved in what is done in its name, and the multifarious ways in which those activities may impact on crime, make Neighbourhood Watch a convenient common-sense peg, but an inappropriate class for the purposes of science (see Laycock and Tilley, 1995). At least that would be the implication of adopting a scientific approach. Science is often slow, while policy and practice will not wait. Science stresses the uncertain, whilst the rhetoric of policy-making asserts confidently and trades on shared understandings with practitioners and the public.

It is worth remembering that the active ingredients in programs are frequently not self-evident. In the case of street lighting upgrades, for example, the alterations in community confidence triggered by them appear to have been capable of effecting changes in crime, rather than the lighting per se (Painter, 1995). In property marking, the publicity showing resident burglars that crime risks were being increased appeared to be the active ingredient in a South Wales experiment, rather than the property marking per se (Laycock, 1985, 1997). In the case of CCTV, the transmitted pictures appear to be but one of a series of mechanisms through which changes in crime levels can be brought about (Phillips, 1999). In the case of mandatory arrest for domestic violence, the changes effected varied according to contrasting mechanisms that were context-dependent. Mixed findings have been reported for the effects in different cities (see Sherman, 1992). These may be explained by variations in reactions to arrest by those living in different social conditions. Arrest is liable

to trigger anger amongst some and shame amongst others. Anger is more likely for those who are unemployed and not attached to stable communities, and shame amongst employed members of stable communities. Anger may precipitate repeat violence and shame inhibit it. Smith et al. (2002) show that many crime prevention initiatives begin to have their effects before they are implemented; the active ingredient is clearly not or not just what is actually put in place. The discovery of diffusion of benefits raises questions about the means by which effects are apparently frequently felt beyond the immediate operational range of measures. Sherman (1990) again furnishes an example in his demonstration of the enduring effects of crackdowns, apparently disrupting crime beyond the time when they are being applied. Each of these examples suggests that surface, lay and policy maker definitions of what comprise key elements in programs can be inadequate and misleading.

The simple association of rate change with a particular measure will be a poor guide to what might usefully be learned and transferred. It is the salient underlying mechanisms triggering the change that are critical: e.g., community confidence, offender uncertainty, heightened perceived risk among offenders, shame, anger and so on. And these more fundamental mechanisms may be triggered in multiple ways by diverse interventions according to context. Evaluations need to tease out the underlying processes to derive practice and policy-relevant lessons. A fixation on surface measures, which may be the overt interest of policy makers, introduces a counterproductive bias overestimating the significance of measures and potentially blinding evaluation studies to the more significant underlying mechanisms. In medicine, the concern with placebos describes a basic effort to sort out physical mechanisms (often tentatively suggested by laboratory and clinical work) from psychological ones. Surface evaluation that simply associates measures with outcomes fails even to make this basic discrimination. We need only to go back to the telephone relay assembly experiments at General Electric's Hawthorne factory to appreciate the difference between direct and indirect mechanisms in producing change (Mayo, 1933; cf., also Smith et al., 2002). There, readers will remember, productivity went up when lighting was increased, but also went up when almost all other changes were introduced. The active ingredient was not the lighting intensity increase per se but the cooperative, self-motivating group brought about by the fact of experimentation. The misleading conclusions that could be drawn for policy without making this distinction are all too obvious. The Hawthorne studies have uncanny echoes in the literature on lighting upgrades and crime (Pease, 1999; Painter, 1995).

Whilst better policy-making and practice may need evaluations that adopt a scientific approach, this will not necessarily initially be popular amongst the policy makers and practitioners we address. Getting good answers entails asking good questions. Good questions are not necessarily the self-evident ones. Colluding with bad questions may win friends in some policy circles. It will not, though, best serve crime prevention and the scientific foundations needed for its improvement.

AN OVERVIEW OF THE FOLLOWING PAPERS

Each of the following papers addresses an important issue in evaluation for crime prevention. They all highlight the potential of well-conducted evaluations to contribute to improvements in policy and practice. They demonstrate the importance of theory both for practice and for evaluation. In particular, they indicate the importance of attending to mechanisms that may produce change or inhibit it. Moreover, they show too a need to be sensitive to the contexts in which crime prevention strategies are developed and measures are put in place. They show that practitioners and policy makers are theory users, and that they can become more effective with better theories and with their better application. They also show how evaluation can help bring theories to the surface, test them and feed findings back either during the lifetime of a project or for future programs. In these ways, the papers published here all constitute efforts to show how scientific evaluation can be conducted in the interest of policy and practice improvement.

Niall Hamilton-Smith takes the knotty issues of displacement of crime and diffusion of crime prevention benefits as unintended consequences of the introduction of crime prevention measures. (Displacement is the relocation of crime from the prevention program project area to neighboring districts. Diffusion occurs when the benefits of a crime prevention program spill over from the project area to neighboring districts.) He treats these as both measurement problems for the evaluator and as practical issues to be addressed by those trying to prevent crime. He shows that the research literature is helpful in shaping our expectations about likely patterns of displacement and diffusion. He also shows that local thinking can usefully be drawn upon in applying this knowledge in specific settings. The implication of his arguments is that in different contexts different patterns of displacement and diffusion of benefits should be expected. Members of the research community therefore need to develop informed hypotheses about potential displacement and diffusion of benefits following the introduction of preventive interventions,

and to adopt bespoke measurements to test the hypotheses. Practitioners need to develop their preventive strategies to minimize unwanted but predictable displacement, and to maximize potential diffusion of benefits. Hamilton-Smith highlights the importance of theory in both practice and evaluation. He shows that the theory needs to focus both on the mechanisms through which displacement and diffusion of benefits are brought about and the particular contexts in which these mechanisms work.

John Roman and Graham Farrell address the relatively underdeveloped but important field of cost-benefit analysis in crime prevention evaluation. They make a trenchant defence of its use in informing policy. They highlight the ubiquity of cost-benefit estimates in decision making over resource allocation at both individual and societal levels, even though these are often implicit or ill-informed. Roman and Farrell show how cost-benefit analysis might usefully be applied broadly to inform crime policy, for example in relation to programs, products, designs and routine practices, both when these do and when they do not have overt crime-related objectives. They show how the demand and supply of criminal opportunities can be modeled, and how to derive the costs and benefits of changes in such opportunities. Roman and Farrell address the tricky issues of non-monetary costs and the means of their estimation, highlighting the problems that follow from excluding such costs. They discuss the possibility of tax-adjustments in respect to the crime-generating activities of manufacturers whose products supply crime opportunities. As with the preceding chapter, Roman and Farrell highlight the importance of attention to mechanisms producing patterns and changes in them, the importance of attention to unintended consequences as well as intended ones, and the use of appropriate methods in the formulation of policy as well as its post-hoc assessment.

John Eck challenges a contemporary methodological shibboleth: that experimental, random controlled trials that are designed to maximize internal validity are always the appropriate "gold standard" (best method) in the evaluation of crime prevention interventions. He highlights the importance of context sensitivity, which is: "the variation in effectiveness caused by implementing the same measure in different social, temporal and physical settings." "The greater the variation," he says, "the greater is the context sensitivity." Eck contrasts small-scale, small-claim interventions with large-scale, large-claim ones, where experimental evaluation may be appropriate. The sorts of place-based, small-scale, small-claim interventions developed in problem-oriented policing are, he suggests, particularly unsuited to classic, technically rigorous experimental methods. What is important in much local small-scale work is analysis of the particulari-

ties of the presenting problem in its context in the light of well established theory — notably that from situational crime prevention and routine activities — and the formulation of a tailored, informed strategy. Orthodox, rigorous, experimental methods aiming to maximize internal validity, he argues, do not make sense here. Instead, simple evaluations examining whether change has occurred and attempting to eliminate obvious alternative factors make more sense. These can help practitioners learn and adapt their interventions. Like other contributors, Eck again stresses the importance of understanding contexts and causal mechanisms in effecting change, and the usefulness of this not only in post-hoc evaluation but also in planning interventions. Eck shows what it is to bring the general findings of science to particular situations for practice, and to learn from the experience of doing so.

Brian English, Rick Cummings and Ralph Straton highlight the different purposes for, types of, approaches to and methods used in evaluations of community crime prevention programs. In particular, action-oriented evaluation is aimed at program development, whilst research-oriented evaluation is aimed at enhancing knowledge. Different time scales and different appropriate methodological standards apply in each case. Methods need to be chosen in the light of purposes. None, they argue, is intrinsically superior. What is actually done will involve trade-offs in the light of the conditions in which the evaluation is conducted. Like Eck, they suggest that attributes of programs themselves are also important in evaluation design. Moreover, as with other contributors, they stress the importance of theory, mechanisms and program contexts in making sense of programs and making decisions about evaluation designs. English et al. sample community crime prevention program evaluations and conclude that there is scope for use of a wider range of approaches. Currently, for example, they find that goal-based evaluations dominate. There is relatively little rich analysis of the ways in which the program is experienced by participants, staff and others implicated in it. This kind of analysis, they suggest, could be helpful in making judgments about how a program is working and ways in which it could usefully be modified.

Dennis Rosenbaum considers the evaluation of multi-agency anti-crime partnerships, which have widely come to be seen as essential for effective crime prevention. He reviews and synthesises a substantial literature from health promotion and drug prevention, as well as crime prevention, to inform the conceptualisation and evaluation of partnerships. The rationale for partnership has to do with the complexity and multidimensionality of the problems they are supposed to address. Partnerships, in theory, promise a way better to define

problems, develop creative solutions, deliver multiple interventions, address problems at multiple levels, and trigger multiple mechanisms that may operate synergistically. Key partnership mechanisms that may deliver these benefits include putting heads together to stimulate new ideas, aligning resources to increase intervention dosage and co-ordination to improve service delivery. Partnerships will vary in what they do and their underlying working theories, informing their "domains of influence," "causal mechanisms," "intervention targets," and "partnership services." Effectiveness may turn on combinations of and interactions between these, but in ways yet to be tested. Rosenbaum provides a range of suggestions for measuring input and context variables, processes involved in planning, program development and implementation, and outcomes. He remains a strong adherent of traditional experimental methods as the gold standard in outcome measurement - as a way of establishing the counterfactual, which is crucial. Rosenbaum also emphasises the benefits of other approaches, however, including case studies, theories of change, time series analysis, qualitative methods, and the use of multiple methods. As with the other contributors to this volume, he strongly emphasises throughout the importance of theory as a guide to measurement and method.



Address correspondence to: Nick Tilley, Department of Social Sciences, Nottingham Trent University, Burton Street, Nottingham NG1 4BU. E-mail: <tilley@home-office.swinternet.co.uk>.

REFERENCES

- Bhaskar, R. (1975). *A Realist Theory of Science*. Brighton, UK: Harvester.
- Clarke, R. (1997). "Introduction." In: R. Clarke (ed.), *Situational Crime Prevention: Successful Case Studies*. Guilderland, NY: Harrow and Heston.
- Eck, J. and W. Spelman (1987). *Solving Problems: Problem-Oriented Policing in Newport News*. Washington, DC: Police Executive Research Forum.
- Eckblom, P. (1990). "Evaluating Crime Prevention: The Management of Uncertainty." In: C. Kemp (ed.), *Current Issues in Criminological Research*. Bristol, UK: Bristol Centre for Criminal Justice.

- (1988). *Getting the Best out of Crime Analysis*. (Crime Prevention Unit Series, Paper No. 10.) London, UK: Home Office.
- and K. Pease (1995). "Evaluating Crime Prevention." In: M. Tonry and D. Farrington (eds.), *Building a Safer Society: Strategic Approaches to Crime Prevention*. (Crime and Justice series, vol. 19.) Chicago, IL: University of Chicago Press.
- Farrington, D. and A. Petrosino (2001). "The Campbell Collaboration Crime and Justice Group." *Annals of the American Academy of Political and Social Science* 578:5-49.
- Goldblatt, P. and C. Lewis (1998). *Reducing Offending: An Assessment of Research Evidence on Ways of Dealing with Offending Behaviour*. (Home Office Research Study No. 187.) London, UK: Home Office.
- Goldstein, H. (1979). "Improving Policing: A Problem-Oriented Approach." *Crime & Delinquency* 25:236-258.
- Laycock, G. (1997). "Operation Identification, or the Power of Publicity?" In: R. Clarke (ed.) *Situational Crime Prevention: Successful Case Studies*. Guilderland, NY: Harrow and Heston.
- (1985). *Property Marking: A Deterrent to Domestic Burglary?* (Crime Prevention Unit Series, Paper No. 3.) London, UK: Home Office.
- and N. Tilley (1995). *Policing and Neighbourhood Watch: Strategic Issues*. (Crime Detection and Prevention Series, Paper No. 60.) London, UK: Home Office.
- Mayo, E. (1933). *The Human Problems of Industrial Civilisation*. New York, NY: MacMillan.
- Painter, K. (1995). "An Evaluation of the Impact of Street Lighting on Crime, Fear of Crime and Quality of Life." Unpublished Ph.D. thesis, Cambridge University.
- Pawson, R. and N. Tilley (1997). *Realistic Evaluation*. London, UK: Sage.
- Pease, K. (1999). "A Review of Street Lighting Evaluations: Crime Reduction Effects." In: K. Painter and N. Tilley (eds.), *Surveillance of Public Space: CCTV, Street Lighting and Crime Prevention*. (Crime Prevention Studies, vol. 10.) Monsey, NY: Criminal Justice Press.
- Phillips, C. (1999). "A Review of CCTV Evaluations: Crime Reduction Effects and Attitudes Towards Its Use." In: K. Painter and N. Tilley (eds.) *Surveillance of Public Space: CCTV, Street Lighting and Crime Prevention*. (Crime Prevention Studies, vol. 10.) Monsey, NY: Criminal Justice Press.
- Scott, M. (2001). *Problem-Oriented Policing: Reflections on the First 20 Years*. Washington, DC: U.S. Department of Justice, Community Oriented Policing Services.
- Sherman, L. (1992). *Policing Domestic Violence*. New York, NY: Free Press.

- (1990). "Police Crackdowns: Initial and Residual Deterrence." In: M. Tonry and N. Morris (eds.), *Crime and Justice: A Review of Research*, vol. 12. Chicago, IL: University of Chicago Press.
- D. Gottfredson, D. Mackenzie, J. Eck, P. Reuter and S. Bushway (1997). *Preventing Crime: What Works, What Doesn't and What's Promising. A Report to the United States Congress*. Washington, DC: U.S. Department of Justice.
- Smith, M., R. Clarke and K. Pease (2002). "Pre-igniting Crime Prevention." In: N. Tilley (ed.), *Analysis for Crime Prevention*. (Crime Prevention Studies, vol. 13.) Monsey, NY: Criminal Justice Press.
- U.K. Her Majesty's Inspectorate of Constabulary (HMIC), (2000). *Calling Time on Crime*. London, UK: author.