THE EVALUATION OF CRIME PREVENTION: PROBLEMS, ISSUES AND CONTEXT

Adapted from Paper for

Journee d'etudes sur l'evaluation des politiques de prevention

Montpellier, France 15 May 1992

Dr Paul Ekblom

Home Office Research and Statistics Directorate

24 April 1997

My first experience of evaluating the impact of a crime prevention scheme was a hard one. As we say in England, I was thrown in at the deep end of the swimming pool. I had to assess the impact on crime of some experimental police patrols aimed at dealing with truants (Ekblom 1979). We developed a careful before, during and after research design, and a range of observational measures of children on the streets and in the shops. Over several weeks, my colleagues and I spent time every schoolday on the frozen streets of Bristol, counting children. There were bad omens from the start. The police sergeant who was driving me round the patrol area for the first time, crashed into the back of another vehicle that was delivering hot meals to elderly people. Bomb scares disrupted the counting of children entering a big store in the city centre. And worst of all, when the week came for which we had planned the truancy patrol, a great snowstorm closed all the schools - so we had rapidly to reorganise our plans. Because of all these unkind events, we could only assess the impact of the patrols with much uncertainty, and to make any useful conclusions we had to rely much more on inference than on direct measures.

Since then, I have conducted several other evaluations, some of which while I worked in the Home Office Crime Prevention Unit (eg Ekblom 1987). Although none has reached the extremes of bad luck I experienced in Bristol, I have learned that evaluation is an extremely difficult task to do well; and that even the most carefully-planned evaluation is very likely to end up messy and complicated. Issues which run right through all evaluations of crime prevention (Ekblom 1990) are **risk** that the evaluation design will fail to deliver conclusive results, or even risk that the preventive action being evaluated will fail to be implemented (Rosenbaum 1986); and **uncertainty** surrounding the conclusions that do emerge. How risk and uncertainty are best handled relates to the **context** of evaluation - the use to which the evaluation results are to be put. These act as the themes that link together the rest of my paper, which in other respects
ranges over a fairly wide field.

I begin with a discussion of the uses and functions of evaluation and move on to consider different types of evaluation. Focusing next on evaluation of impact, I examine some technical difficulties which illustrate the significance of risk, uncertainty and context, and consider how these themes relate to different sets of users of evaluation results. To match the shortcomings on the part of users, I then identify shortcomings on the part of evaluators themselves and make some suggestions about how their conduct could be improved to suit the context in which they are conducting particular studies. I then finish by raising the issue of how far it is possible to teach crime prevention practitioners to evaluate their own work.

Uses and functions of evaluation

At this point, if I could be a little didactic, it may help if I set out a modest framework distinguishing between several different uses and functions of evaluation (Figure 1). At its simplest, evaluation can serve a retrospective, **accountancy**-type function - merely ensuring that money was properly and efficiently spent. It can also serve a **research** function, with findings contributing to the rigorous testing of theories and development of social scientific knowledge. Finally, it can serve a prospective function, **guiding decisions** in the real world about policy and practice. The **scale** of these decisions can vary from local to national, and their **scope** can vary from continuance (should this one scheme continue?), to expansion (should we put more money into the scheme?) to reproduction elsewhere (should we set up more schemes like this one?).

Coming as I do from a government department, I am of course mainly interested in the prospective, decision-guiding function. In the crime prevention context, this kind of evaluation is one component in what has become known as the **preventive process** (Ekblom 1988), widely-applied in the British and North American context (Clarke 1992), with some success. This is a rational, information-based, 'problem-oriented' approach (Goldstein 1979, 1990) which can be used by the police or local crime prevention organisations, to identify individual, local and specific crime problems and to select the most appropriate solution to those problems. It has five stages:

* Identification of crime problem (including the technique of crime pattern analysis)

* Development of solutions to problem

* Implementation

* Evaluation

To give an example of the preventive process in action, in the early 80s there was a wave of robberies of small post offices in London (Ekblom 1987). The Post Office Investigation Department applied the technique of crime pattern analysis (Ekblom 1988) to the records of the robbery incidents, and in particular examined the methods of offending used by the robbers. This revealed that many robbers were exploiting physical weaknesses in the anti-bandit screen, and in security procedures such as closure of doors. Both were improved through a major preventive programme, and the best estimate the data could allow was that

the security initiative reduced the robberies by about one third from their peak level. This appeared cost-effective, and a good return on investment.

Types of evaluation

Evaluation is not, of course, a single homogeneous activity. **Within** the global concept of evaluation, the major distinction to make is between a focus on **implementation** and a focus on **impact** or effectiveness. On the implementation side, as Figure 2 shows, we have managerial-type **monitoring** (was the preventive scheme targeted on the right people, places or property? Did the scheme deliver appropriate activity to the targets?) and the study of **process** (what were the practical difficulties and problematic issues raised in implementation?). On the impact side, on which I shall concentrate in the rest of my talk, there are a number of key questions which all prospective impact evaluations of crime prevention have to ask. These questions are:

i) Was there a distinct fall in the crime level associated with the preventive action, as opposed to random fluctuation and the progressive influence of background trends?

ii) If yes, what proportion of the observed fall can be attributed to the preventive action, as opposed to all the other causal influences upon crime levels that are usually present?

iii) What side-effects have there been? Fear of crime can sometimes be increased by a badly-designed crime prevention campaign. There can also be displacement or the opposite, diffusion of benefit.

iv) How cost-effective was the preventive action?

v) Can the action be reproduced elsewhere? Were there any 'hidden ingredients' in the success?

Technical difficulties in evaluating impact

It is easy to pose these questions, which are not especially technical, but to answer them, of course, can be extremely difficult and **very** technical. Let me give one or two illustrations.

On question i), the main problem in detecting change and estimating its size is random fluctuation. In crime statistics at the national level or the city level, aggregation over large areas and a whole year often gives the appearance of a deluge of offending. But in the shorter term, and at the very local level - where current approaches to crime prevention focus - crimes are rare events and fluctuation can be severe. This was brought home to me on my very first ride in a police patrol car, in a place in the North-East of England called, appropriately enough, Pity Me. The police beat I was assigned to was supposed to be high in crime, but as the hours of darkness crawled by and nothing happened, my escort kept shaking his head and

saying 'you should have been here last week!' Since then, I have met the YSHBHLW phenomenon on many occasions, to my weary frustration.

One way to filter out such random fluctuations (together with seasonal changes and background trends) is to use the statistical technique of time series analysis. Unfortunately, it is often difficult to obtain a continuous series of crime data that is sufficiently long (eg 60 months' worth of figures) - the police sometimes destroy data after 2 years, beat boundaries are changed and so on. In any case, background trends in crime may not be stable and linear - they may show exponential growth or decline, and therefore may be difficult to estimate and then to subtract in order to leave behind a clear picture of relative change (for an example, see Ekblom 1987).

Paradoxically, smaller preventive schemes may have more impact on crime than larger ones, but proving this may be harder because of greater interference from random fluctuation. Bennett (1990) fell foul of this in his evaluation of neighbourhood watch schemes in London. He chose two very large schemes in order to minimise the problems of random fluctuation - but unfortunately their very size also meant they were not very well implemented, so the test was less fair than it might have been.

On question ii), keeping track of **cause and effect** is extremely difficult. There may be many other factors at work influencing the level of crime in the experimental area - police patrolling, the local economy, the weather even. All conspire to confound even the most sophisticated research design. Even the displacement of crime can mislead, if it shunts crime from the experimental area to a control area, giving a falsely-increased estimate of the size of impact. Inference here is very difficult and the evaluator often has to **explore** the data, and make professional judgements about the plausibility of different explanations for an observed fall in crime. Usually it is desirable to aid this process by developing theoretical models of the processes which are supposed to be happening. *This important aspect of evaluation has been developed since this paper was written in 1992, through the Scientific Realist approach of Nick Tilley - focusing on the causal mechanisms by which particular preventive effects are triggered in particular contexts (eg see PRG Crime Prevention series 42, 47.). Specifying the mechanisms by which a preventive scheme could work/may have worked is vital in a) ensuring there is something worthwhile evaluating in the first place rather than merely woolly ideas, b) helping obtain clear answers from the subsequent evaluation, including learning from failures as well as successes and c) understanding what aspects of a successful scheme may be reproduced in what other contexts.*

Taken together, it is rarely possible for the evaluator to draw a 'watertight' conclusion about the impact of preventive action. There is always a degree of uncertainty, which can be very large and impossible to quantify. To return to a theme I introduced earlier, there is also an element of risk in conducting any evaluation - it is rare for things to work out as originally planned. Well-chosen control areas can show unexpected and unexplained falls in crime (eg Bennett 1990); the police can suddenly increase patrols in the experimental area (eg Webb and Laycock 1992) and so on.

Reducing risk and uncertainty

Reducing the risk of 'measurement failure' (Rosenbaum 1986) - that is, the failure of an evaluation to detect the real effect of a successful crime prevention scheme - is costly. So is reducing the margin of uncertainty that surrounds the conclusions of any evaluation. In both cases, collecting more data, using more than just one or two experimental and control areas, applying more sophisticated designs and analytical techniques, and allowing more time for the preventive activity to take effect and to be reliably measured, will help minimise these problems - but they will never reduce them entirely. There is an inescapable tradeoff between resources put into the evaluation, the quality and certainty of the results of the evaluation, and timing - how soon those results can be available to guide decisions. Depending on circumstances and the function of the evaluation (as discussed earlier and shown in Figure 1), the most appropriate balance to be struck here will be very different. In one context, for example, the local police will want to know whether a particular scheme against car park crime has worked at one location. Other, more senior police managers may want to know whether the experimental scheme is worth reproducing elsewhere; policy-makers in central government may want to know if the scheme is worth reproducing nationwide, and whether preventing crime in car parks is likely to have a significant effect on car crime as a whole. As we go up the scale, it makes increasing sense to change the parameters of the evaluation: to reduce its uncertainty and risk by investing in more resources and greater sophistication of design, and taking a longer time over it.

The different users of evaluations

There are, of course, several different parties wishing to use evaluation results: academics, local and national administrators and policy-makers, commercial companies and practitioners such as the police or local crime prevention teams. Their understanding of evaluation, its various functions, its constraints and possibilities, will differ. So, too, will their interest in the results - sometimes quite dramatically. Any professional evaluator has to be fully aware of these considerations raised by the context of utilisation of evaluations, so I will give several examples. As might be expected, many of the differences of opinion that arise centre on the issue of uncertainty and how to handle it.

Practitioners

Many practitioners of crime prevention, such as local police Crime Prevention Officers or crime prevention coordinators still have rather a limited understanding of evaluation - particularly concerning statistical inference and cause and effect (for example, they might triumphantly proclaim 'a 50% fall in crime' on the basis of a fall from two burglaries per month to one; or neglect to take account of the introduction of parking charges at a car park shortly before the preventive measures were set in hand). Moreover, they will understandably wish to view the fruits of their hard work in the best possible light. So it is unsurprising that, on several occasions, I have seen a head-on confrontation between evaluators and evaluated when negative findings are reported (in one case, the evaluator was nearly lynched).

Administrators and policy-makers

In recent years, the British Government has accorded increasing importance to good management of its operations and those of the public services for which it is responsible,

particularly financial management and the securing of 'value for money'. Evaluation obviously has a central role to play here, but Government administrators and policy-makers, for their part, have not yet fully come to terms with this. Previously, evaluators have been able to 'parachute in' and conduct an evaluation on their own terms. Now, however, they are increasingly 'shackled together' with their administrative colleagues who - of course! - want an answer that is reliable, valid, fast, cheap and risk-free. Difficulties often emerge when evaluators tell them that they cannot have speed, reliability and economy all at the same time. In some cases, it is all too easy for administrators to maintain unrealistic expectations about what evaluation can achieve - and to try to leave the resolution of such tradeoff questions to the professional evaluators alone.

Treasury

In Britain there also seems to be a need for Treasury to adapt to the constraints of professional evaluation. Treasury has very laudably begun to require administrators in the 'spending' departments of government to evaluate their projects and programmes. However, Treasury itself has unfortunately not yet moved very far beyond an understanding of evaluation as retrospective accountancy. A Treasury booklet on 'how to evaluate policy' is full of important **fiscal** measures, but contains less than half a paragraph on what I refer to as professional evaluation issues such as research design, and is virtually blind to the kind of tradeoffs and constraints on evaluation that I have been discussing. The Canadian government seems rather ahead of our own in its understanding of evaluation (perhaps this is due to the French influence), and I commend its publications to you (Treasury Board of Canada/Comptroller General 1981a, b).

Shortcomings on the part of evaluators

So far, I have pointed out some limitations of the understanding of evaluation of practitioners and administrators - but to be fair, there have been some serious shortcomings on the part of some evaluators too. Coming, as most evaluators do, from a background in social research, they may fail to realise that they have changed the context in which they are operating, and consequently fail to realise that the function of the evaluations that they are conducting have also changed. 'Academic' evaluations, and academic research require, of course, exceptionally rigorous quality control, in order to ensure that only the most reliable and valid findings are admitted to the body of cumulative academic knowledge. In particular, the risk of **mistakenly inferring a successful impact** on crime is considered more important to avoid than the risk of **failure to detect an effect**. This is embodied in the usual tests of statistical significance.

This orientation can be entirely inappropriate when rigidly applied to evaluation in its **prospective** function of guiding decisions in the real world. Practitioners and administrators cannot take 'don't know' for an answer merely because p has just exceeded .05, a value which was set by convention for reasons of quality control. Nor may they always give greater importance to the avoidance of 'false positive' errors than the avoidance of 'false negatives'. It depends on the consequences, to the users, of making one or other type of error regarding the decision in question. What users do want is the clearest answer possible for the resources

and time available - and increasingly, they want a **quantified** answer. To what **extent** does their scheme work, at what **cost**?

The conduct of professional evaluators

Social researchers, when we become professional evaluators contributing information for the guidance of decisions in policy and practice, have therefore to learn to adapt our methods to circumstances. We have to develop our methodology to give the best possible answers to questions posed in a real-world context of financial constraint, short timescales, deficient data and imperfect experimental control. Do not misunderstand me: I am not remotely advocating a relaxation of professional discipline. Rather, I am suggesting that professional evaluators have to meet more exacting requirements than our colleagues in academic research. Instead of working within comfortably fixed parameters such as 'p equals .05', we have to work from first principles. In some cases, the .05 criterion has to be relaxed; but in others, it might actually have to be **tightened** to .01 or better.

More generally speaking, we should learn to be versatile in the use of statistical techniques (and aware of their individual limitations); clever and imaginative at creating research designs, able to identify weaknesses imposed by external constraints and to find the next best remedies; accustomed to collecting and using diagnostic information, much of which may be qualitative; and able to strike sensible balances between different sources of uncertainty - for example it is little use deciding to devote a great deal of resources to reducing statistical uncertainty, if you can drive a coach and horses through the cause-effect uncertainty. Evaluators must be prepared to adjust the parameters of the study, its resources and timing according to the context in which it is to be used. In some circumstances an expensive, very high-quality evaluation will be appropriate; in others, a simple, cheap, rough and ready one: 'Rolls-Royce' versus 'Deux Chevaux' if you like (although I am not implying English superiority here!) But in **all** cases, to be able to decide **which** is needed, the evaluator requires a sophisticated grasp of first principles and an awareness of the functional framework mentioned above (Figure 1). This has particular implications for the conduct of researchers at the start and the finish of evaluations.

At the **start**, the evaluator must be prepared to raise the issue of uncertainty with the user of the evaluation, pointing out in particular the trade-off between false positive and false negative errors. Rather than allowing administrators to delegate the task to the evaluators, we evaluators must stand our ground and make the setting of cost, time, uncertainty and risk parameters a **joint** responsibility. But the tradeoff is not infinitely elastic: there will be times when the evaluator should be prepared to tell the user that within the constraints of resources and timing that have been set, the uncertainty is likely to be so great that an evaluation would be worthless. Having, let us assume, decided to go ahead with the evaluation, the evaluator should create a design within the parameters set.

At the **end** of an evaluation, the analysis of data will have to be exploratory - it is really only at this stage that some of the problems of data or of inference emerge. Having finally produced some results, the evaluator cannot simply hand over an unqualified answer and run the risk that it will be taken at face value by the user. Once quoted, numbers have a tendency to forget their uncertain origins. Fortunately, the competent manager or administrator is

accustomed to decision-making under conditions of uncertainty, so the solution is to provide the user with an assessment which contains the evaluative judgement together with the associated range of uncertainty. On the quantitative side, this is done by providing an estimate of the size of a preventive effect together with **confidence limits**. (These confidence limits need not be completely based on statistical tests, which only cover some of the sources of uncertainty - one might give a range of values for the preventive effect based on a comparison of estimates derived from different measures (eg Ekblom 1987).)

To take a broad view, there is, I think, a need for a sustained period of mutual education and accommodation between administrators and evaluators, much like that which has successfully happened in Britain over the last 15 years regarding administrators and social researchers. Such a process could be accelerated by the adoption of a practice of drawing up explicit contracts between the two parties, which set the broad parameters for each individual evaluation.

Do all evaluators have to be professional evaluators?

As the new approaches to crime prevention move from experimental activity to routine, there is a desperate need for reasonably-well-evaluated information on 'what preventive methods work, in what circumstances', in order to guide practitioners in their choice of preventive strategies. Unfortunately, there is so far only a rather limited body of this knowledge (see Clarke 1992, for examples). One way to reduce this 'knowledge gap' is to increase the number of professionally-trained evaluators. This is obviously good for our employment prospects, but there are limits. A cheaper approach might be to try to teach some of the skills of evaluation to lay practitioners in crime prevention; in particular to give them cheap and simple evaluation techniques that will serve them adequately for most of the time. This sounds efficient and pleasantly egalitarian, and various attempts are being made to carry the idea forwards, in the Home Office and elsewhere. But personally speaking, I have rather strong doubts about how far one can proceed in this direction. The possibilities for making the wrong inferences in an evaluation are enormous - and the more limited the data and the design of the evaluation, the greater the scope for making such mistakes. We may not be talking about making a conclusion '80% as good as that provided by a professional evaluator, for 20% of the cost': errors in evaluation tend not merely to gradually degrade the quality of the result - they can lead to completely the wrong answer. Paradoxically, saving money and/or time on the design and data collection aspects of an evaluation means greater sophistication is often required to extract a reliable and valid conclusion to guide decision-making.

Professional judgement, in my opinion, is most needed at the start and finish of an evaluation, tying in with what I said in the previous section. First, in deciding whether the decision-making context and the preventive activity to be evaluated necessitate a simple or a sophisticated design; and second, when the design is under way, in extracting and interpreting the pattern of results. Knowledge of techniques may be less important than the possession of a clear conceptual framework covering the key questions evaluation asks, plus a list of commonly occurring alternative explanations for findings (Campbell and Stanley's (1966) 'threats to internal validity') which must be tested and eliminated. It may be that a **combination** approach is best, with crime prevention practitioners given elementary guidance and training in evaluation, together with access to a professional advisory and

support team able to help them at critical moments of designing, conducting and analysing the results of an evaluation (see Berry and Carter 1992). But ultimately, the answer to this important speculation may be to try it and see - to set up a range of evaluation exercises and pit professionals against lay evaluators acting alone or with some guidance.

To finish, I have raised a lot of difficulties for the would-be evaluator - but I hope you take them less as a discouragement than a **challenge**. And there is one happy ending to report - the police sergeant who crashed his car in Bristol is now a chief superintendent and very high up in crime prevention! *Former dep head of HOCPC, now retired and working for a security company...*

These points are all made, in more academic style and more detail, in Ekblom, P. and Pease, K. (1995). *'Evaluating crime prevention.'* In Tonry, M. and Farrington, D. (Eds.), **Building a Safer Society: Strategic Approaches to Crime Prevention**. **Crime and Justice: A Review of Research**, vol. 19. London and Chicago, University of Chicago Press, pp. 585-662. *Copy available from Paul Ekblom*

REFERENCES

Bennett, T (1990). *Evaluating Neighbourhood Watch*. Aldershot, Hants: Gower.

Berry, G and Carter, M (1992). *Assessing Crime Prevention Initiatives: the first steps.* CPU Paper 31. London: Home Office.

Campbell, D and Stanley, J (1966). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Clarke, R (Ed) (1992). *Situational Crime Prevention: successful case studies*. New York: Harrow and Heston.

Ekblom, P (1979). 'Police truancy patrols' in Burrows, J, Ekblom, P and Heal, K *Crime Prevention and the Police*. Home Office Research Study 55. London: HMSO.

Ekblom, P (1986b). *The Prevention of Shop Theft: an approach through crime analysis*. Home Office Crime Prevention Unit Paper 5. London: Home Office.

Ekblom, P (1987). *Preventing Robberies at Sub-Post Offices: an evaluation of a security initiative*. Home Office Crime Prevention Unit Paper 9. London: Home Office.

Ekblom, P (1988). *Getting the Best out of Crime Analysis*. Crime Prevention Unit Paper 10. London: Home Office.

Ekblom, P (1990). 'Evaluating crime prevention: the management of uncertainty.' In Kemp, C (Ed), *Current Issues in Criminological Research*. Bristol: Bristol Centre for Criminal Justice.

Ekblom, P (1992, in press) 'The Safer Cities Programme Impact Evaluation: problems and progress' in *Studies on Crime and Crime Prevention* 1. Stockholm, Sweden: National Council for Crime Prevention/Scandinavian University Press.

Goldstein, H (1979). 'Improving policing: a problem-oriented approach'. *Crime and Delinquency* (April): 234-258.

Goldstein, H (1990). *Problem-Oriented-Policing*. New York: McGraw Hill.

Home Office (1991) *Safer Cities Progress Report, 1990-1991*. London: Home Office.

Rosenbaum, D (Ed) (1986). *Community Crime Prevention: does it work?* London: Sage.

Treasury Board of Canada/Comptroller General (1981a). *Guide on the Program Evaluation Function*. Ottawa: Treasury Board of Canada.

Treasury Board of Canada/Comptroller General (1981b). *Principles for the Evaluation of Programs by Federal Departments and Agencies*. Ottawa: Treasury Board of Canada.

Webb, B and Laycock, G (1992b). *Reducing Crime on the London Underground: an evaluation of three pilot projects*. Home Office Crime Prevention Unit Paper 30. London: Home Office.