

EVALUATION OF SAFETY AND CRIME PREVENTION POLICIES

England and Wales

Tim Hope¹

June 2008

Workshop on the Evaluation of Safety and Crime Prevention Policies
Bologna, 10-12 July 2008

Work-package 7 (Methodology and Good Practices)
CRIMPREV – Assessing Deviance, Crime and Prevention in Europe

Citation:

Hope, T. (2009) 'Evaluation of safety and crime prevention policies in England and Wales' In Ph. Robert (Ed.) *Evaluating Safety and Crime Prevention Policies in Europe*. Brussels: VUBPRESS Bussels University Press. 91 – 121. ISBN 978 90 5487 588 8; pp. 91-121

¹ Professor of Criminology, Keele University and Senior Visiting Research Fellow, Scottish Centre for Crime and Justice Research (CJ-QUEST), University of Edinburgh.

Introduction

This paper is concerned not so much with the substantive results of evaluation research concerning safety and crime prevention policies as with the question of the *evaluability* of community crime prevention, using examples from studies carried out in England and Wales. The concept of evaluability refers to *those aspects of the institutional context of an intervention that affect the methodology of evaluation applied, and the validity of inferences drawn from applying that methodology*. Unlike the discussion of methodology *per se*, the analysis of evaluability is a reflexive approach to criticism that seeks to give an account of how institutional context (the circumstances in which a particular evaluation is carried out) is related to the particular methodological approach adopted. The reflexive study of evaluability is far less developed than the study of evaluation method itself, yet the governance issues involved in evaluation research often have a crucial influence on its design, conduct and findings. Just as contemporary British governments have invested heavily in the evaluation of community crime prevention, so governance issues have had a critical impact on their evaluability.

Why evaluate? - The scientific case

From the social scientific point of view, the classic statement about the purpose of evaluation research is Campbell (1969). Drawing upon the ethos of the 'Great Society' programmes of 1960s America, Campbell argued that an 'experimenting society' would be a 'good society' because social and political learning would be advanced by the application of scientific method. The use of scientific criteria to inform political choice would consequently lead to progress in the social policy sphere just as it seems to do in the field of medical science. Nevertheless, Campbell also saw the need to establish an experimental ethos in public administration for attaining this goal. The chief obstacle to establishing this ethos was not the absence of knowledge to experiment with, nor even the practical difficulties of implementing scientific methods of evaluation (which are the subject of evaluation research methodology), but the political obstacles in the way of informed social learning and choice stemming from the nature of governance itself. As well as placing science as a methodological tool for governments to use, in Campbell's vision of the reform society, science would also need to be a methodological tool with which to hold government accountable for its actions to a wider public interest. Without an external (publicly-accessible) method of accounting for political choice, public administrators have built-in incentives not only to avoid critical scrutiny of their public record and achievements but also for making political capital by claiming success for what they have purportedly done. Thus, the rigorous application of objective scientific method aspires to raise evaluation research above politics; providing evidence of What Works on a supposedly less partial and partisan basis than that of political discourse and competition.

Why evaluate? - The political case

Usually in the UK, the type of community crime prevention intervention, where and on what it should be targeted, and who should benefit, are decided in advance by the sponsors. There are very few examples of 'spontaneous' community self-help and

self-organisation. Aside from the academic study of ‘social movements’ (covering spontaneous community action), evaluation research is itself an applied research activity intimately tied-up with the process of government. Community crime prevention programmes are *interventions* by government into civil society and it is the nature of these interventions, and their sponsors’ needs, that affect their evaluability. In the UK, the sponsors of community crime prevention are usually the national government, and the purpose of these interventions is to implement its policies². Thus, in answer to the question “why evaluate?” there are three motives or purposes that government sponsors have in commissioning or engaging with evaluation research:

1. *Accountability* – the view that a government’s actions should be subject to some means of scrutiny accessible to those who elect it. Social scientific research methods are thus regarded as a superior means for ensuring political accountability.
2. *Technology* – the view that the evaluation of governmental programmes will reveal information that will provide a ‘blue-print’ for what to do and how to do it. Social scientific research methods are thus regarded as a superior means for developing a technology of governance that will help with the engineering of social reform – that is, to *discover* ‘what works’.
3. *Validation* – the view (often a political perception) that the purpose of evaluation is to provide proof of the desirability of a policy (and its sponsors) and/or its superiority vis-à-vis that of policy (and political) alternatives – that is, to *demonstrate* ‘what works’.

Since evaluation is a political process, these three ‘political’ considerations have each played a crucial role in the conduct of evaluation research on community crime prevention. Distinctions in British policy and practice between them have become blurred and confused (Hope, 2002a), sometimes intentionally (Hope, 2004).

² This pattern tends to reflect the form of public administration in The United Kingdom. The UK is a unitary state, whose central government is accountable to the UK Parliament, consisting of three separate jurisdictions – England and Wales (containing the great majority of the population of the UK), Scotland, and Northern Ireland. Scotland, Wales and Northern Ireland possess their own devolved governments, though their statutory powers vary considerably both amongst themselves and vis-à-vis the UK government. Sub-nationally, there are separate, directly elected local governments and non-elected public authorities (including separate police authorities) that have limited tax-raising powers but cannot act *ultra vires* of national legislation. In all cases, the great majority of revenue required to fund local public services is collected by central government, and is subsequently disbursed to local administrations, with varying degrees of local government responsibility and central government fiscal control. In recent years, central government expenditure has been tied to public service performance regimes set by legislation (Hope, 2002).

Part 1: the evaluation of safety and crime prevention policy in England and Wales

This paper is concerned with research in England and Wales aimed at evaluating community crime prevention, which is defined as *interventions aimed at changing the social conditions that are believed to sustain crime and insecurity in common environments*, most often residential communities (Hope, 1995). Specifically, interventions often aim to change the institutional structure of local communities in ways that increase the exercise of social control both by private citizens and the authorities (Hope and Karstedt, 2003; Hope, 1998; 1995; Foster and Hope, 1993). ‘Institutions’ and ‘social control’ can be defined as both informal and formal – examples of which are given in Figure 1. In essence, these are interventions that address residential places (neighbourhoods) as holistic entities (communities) which encapsulate their own internal social relations and dynamics, including the interaction of formal and informal institutions to ‘co-produce’ safety in common environments (Hope, 1995). As such, they are *Comprehensive Community Initiatives – CCIs* (Connell and Kubisch, 1998).

Figure 1: Dimensions of community crime prevention

	Formal	Informal
Institutions	Police Government	Families, networks Community associations
Social Control	Law enforcement Environmental design Situational crime prevention	Natural surveillance Invocation of authority Informal social control Socialisation

Types of programmes

There have been two generic kinds of central government programme with potential to affect community safety:

1. *Crime Prevention*. Large-scale programmes in England and Wales have included the *Safer Cities Programmes* and the *Crime Reduction Programme*, along with funds available through regional programmes to fund local authority projects, since 1998 via the local government machinery of *Crime and Disorder Reduction Partnerships* (Hope, 2005, 2002). These programmes have included the range of activities described in Figure 1. Notable investments have been in *Neighbourhood Watch*, public *Closed-Circuit Television (CCTV)*, and *Street-Lighting*. There has also been a steady investment in community policing, including *Problem-Oriented Policing*, *Reassurance Policing* and, most recently, *Neighbourhood Policing*. In most of these funding programmes, the projects have been small-scale

‘demonstration projects’ targeted on local neighbourhoods. In general, authorities seek to *deliver* specific crime prevention services to communities. Sometimes these are evaluated for cost-effectiveness, and sometimes they have been accompanied, or comprise part of, national programme-level evaluations. A common governance theme is partnership - between central and local government, between governmental agencies (e.g. local government and police services), and between these bodies and civil society (whether private citizens or corporate bodies). Nevertheless, there are also power structures, hierarchies and organisational cultures amongst agents that affect how they interact together ‘in partnership’ and how and what outcomes are produced.

2. *Urban Regeneration*. Community safety has also been one of the objectives of urban regeneration programmes, though it has not been an exclusive or sole objective. In contrast to the direct effects of programmes delivering specific crime prevention, regeneration programmes tend to address community conditions that give rise to crime - including those of social exclusion and community empowerment – thereby setting in train community-level change that leads to greater social control (Hope, 2005). From the 1970s onwards, regeneration programmes have become increasingly focussed and specific, moving away from main-stream funding towards more targeted demonstration programmes. National programmes have included: the *Priority Estates Project* (PEP), the *Design Improvement Controlled Experiment* (DICE), the *Single Regeneration Budget* programme (SRB), and most recently, the *New Deal for Communities* (NDC) and *Housing Market Renewal* (HMR) programmes. During the course of this development, community safety has become an increasingly salient, and distinctive, part of the package of issues to be addressed.

Evaluation research into community crime prevention

The history of evaluation research in community crime prevention mirrors the development of crime prevention policy and practice in the UK. Key evaluation studies have included³:

1. *The Home Office Crime Prevention Feasibility Study*. This was aimed at demonstrating the application of systematic (evidence-based) prevention and multi-agency partnership working (Hope and Murphy, 1983; Gladstone, 1980). The research utilised an action-research, case-study methodology.
2. *Neighbourhood Watch*. An evaluation of the impact of a pilot Neighbourhood Watch programme in two residential neighbourhoods, utilising a quasi-experimental research design (Bennett, 1990). This study also introduced the use of social survey data as a source of measures of impact (see also studies 3, 4, and 6).
3. *Community Policing (Fear Reduction)*. A quasi-experimental evaluation of a policing initiative aiming to reduce fear of crime and improve the quality of life of residents on two residential estates (Bennett, 1991).

³ This is a selective rather than comprehensive list, selected to represent key issues or developments in evaluation methodology, discussed in Part 2.

4. *Priority Estates Project Evaluation Study*. A quasi-experimental evaluation of the effect on community safety of a housing environmental and management (urban regeneration) programme on two estates in two different English cities (Foster and Hope, 1993).
5. *The Kirkholt Burglary Reduction Project (Repeat Victimisation)*. A multi-agency initiative, including situational, offender-based, and social prevention measures (Forrester et al., 1990; 1988). It has since become famous for pioneering the prevention of repeat victimization and for being considered an early exemplar of 'realist' evaluation (Pawson and Tilley, 1997). The Kirkholt Project has spawned a number of evaluations in the UK and elsewhere focusing on repeat victimization prevention (Farrell, 2005).
6. *The Safer Cities Programme*. A multi-agency, multiplex, multi-site crime prevention programme. Phase 1 ran from 1988-1995, funded and managed by the Home Office; (Phase 2 ran between 1994 and 1998 funded and managed by the ministry for local government). The Programme itself was monitored administratively (Knox *et al.*, 2000). A key focus of Phase 1 was the prevention of burglary: using Geographic Information Systems (GIS) technology and micro-econometric modelling techniques, Ekblom *et al.* (1996), carried out probably the most statistically sophisticated study in an effort to measure programme intensity and incorporate selection bias (see below). Other, smaller scale evaluation studies also focused on burglary, though using mainly qualitative techniques within a 'realist' approach (Tilley 1993; Tilley and Webb, 1994).
7. *Design Improvement Controlled Experiment (DICE)*. Carried out between 1989 and 1997, DICE consisted of radical design improvement schemes on selected council housing estates in England, the evaluation used mixed methods (including surveys) to evaluate the impact, costs and benefits of the schemes. (<http://www.communities.gov.uk/archived/general-content/citiesandregions/221449/>).
8. *The Crime Reduction Programme (CRP), 1999-2002*. This was the largest programme of community-based crime prevention ever attempted in the UK, with an original budget of £250 million (315 million euros, approx.), allocated mainly to local projects, focusing upon different, thematic, crime problems, with 10 per cent ear-marked for independent evaluation studies of cost-effectiveness (Maguire, 2004). Apparently, more than 80 evaluation reports were produced, though fewer than that have been published (Hope, 2008). The most intensively evaluated component of the CRP was the *Reducing Burglary Initiative – Phase 1* (Hope, et al., 2004; Hirschfield 2004; Millie and Hough, 2004; Bowles and Pradiptyo, 2004; Kodz *et al.*, 2004). The CRP also contained a programme of around £150-£200 million (189 – 252 million euros, approx.) for *CCTV* systems in local areas, which has also been evaluated (Gill and Spriggs, 2005).
9. *Community Policing (Reassurance Policing/Neighbourhood Policing)*. "Reassurance Policing is a model of neighbourhood policing which seeks to improve public confidence in policing by involving local communities in identifying priority crime and disorder issues in their neighbourhood which they then tackle

together with the police and other public services and partners” (<http://www.crimereduction.homeoffice.gov.uk/policing17.htm>). There was evaluation of the outcomes of the National Reassurance Policing Programme (NRPP) in England between 2003/04 and 2004/05 (Tuffin *et al.*, 2006). “...The implementation of *neighbourhood policing* has represented a significant undertaking for the Government, police service and their partners. Neighbourhood policing was initially piloted at a ward level as part of the [NRPP]. Following [this] the three-year Neighbourhood Policing Programme (NPP) was officially launched in April 2005, and has sought to deliver on the Government’s commitment for every neighbourhood in England and Wales to have a neighbourhood policing team by 2008” (Quinton and Morris, 2008: iv); an early-stages evaluation report is available (*ibid.*).

Part 2: the evaluability of community crime prevention

Evidence-Based Policy and Practice?

British central governments have invested heavily in crime prevention over the past Quarter-Century. Since 1997, ‘New Labour’ Governments have specifically coupled crime prevention (or, as they call it, *Crime and Disorder Reduction*) with the ‘ideological practice’ of *Evidence-based Policy and Practice* (EBPP), encapsulating the policy pursuit of ‘What Works’ (Goldblatt and Lewis, 1998). The public administration of EBPP is to be a matter, firstly, of accumulating expert knowledge and then having that expertise adopted and applied more generally. The ideal model is of an influential and credible government, guided by expert knowledge, ‘demonstrating’ by practical example that certain programs ‘work’ and thereby persuading others to adopt them (Hope and Karstedt, 2003). Tilley and Laycock (2000), for example, describes the research and implementation programmes involved in the British Government’s development and dissemination of ‘evidence-based’ crime prevention policy. Increasingly, the results of evaluations are to be aggregated and assessed by *systematic review*, whose role is to act as a filter and quality-control for implementing interventions, based upon methodological criteria concerning the ‘quality’ of the evidence available from completed evaluation studies. One such set of criteria – the Scientific Methods Scale (SMS) (Farrington, 2003) – has become increasingly influential as an ostensible criterion for central government evaluation research (Hollin, 2008; Hope, 2005c).

The adoption of community crime prevention practices locally is also a continuation of project-based demonstration. Thus, once persuaded, those with the authority to act are enjoined further to establish and evaluate their own ‘projects’ so that they too can demonstrate benefits, both to their constituents as the recipients of prevention, and to the public purse, that cost-effectiveness has been achieved. The mode in which prevention projects are enjoined to operate conforms, in most respects, to a model of rational strategic planning (Hope and Karstedt, 2003; Hope and Sparks, 2000). Knowledge of a particular crime problem is to be ascertained in advance of trying to implement action, and the selection of services and measures to be delivered is then tailored to the key features of crime manifest in the specific target area or group. The model of EBPP combines the governance precepts of the ‘New Public Management (NPM)’ (Hope, 2002b) with prudential governmental action against social risk (O’Malley, 2000). The method of demonstration by practical example is believed to

work because those who have the responsibility and authority to take action will be shown how to identify prospective benefits accruing in their self-interest, as articulators of the interests of their constituents or ‘customers’. They will be thus persuaded to take requisite action to maximise the utility of their policies on the basis of a calculation of the cost-effectiveness of individual preventive interventions (e.g. Sherman et al. 2002). Syntheses of the evidence from evaluation research are used to develop practical advice, guidance and ‘tool-kits’ for local practitioners (e.g. <http://www.crimereduction.homeoffice.gov.uk/>).

The practical problems of rigorous social experimentation are well-known – such ‘quasi-experiments’ may be difficult to set-up, individual experiments may become un-representative, or may involve unethical choices affecting ‘subjects’, the random administration of penalties, and the denial of beneficial treatments. The problem for policy-makers, though, is a different one – the *Martinson Problem* (Pawson and Tilley, 1997) or, in other words, that programmes may not always be judged to be successful despite the ambitions of their sponsors. For the policy-maker, the ‘threat’ posed by evaluation research is also different than for the social scientist. Social science is oriented towards the avoidance of *Type I Error* – that is, to avoid concluding that a finding is true when it is not. For the policy-maker, however, the situation is reversed: although policy-makers who make Type I errors may well incur opportunity costs to the wider public interest (i.e. by diverting resources towards ineffectual programmes and away from desirable policies), the direct *political* costs of making alternative *Type II Errors* (i.e. to conclude that a finding is false when it is true) are much greater – risking throwing away their programmes and their credibility needlessly. Thus, evaluation research always contains the potential to threaten the political capital vested in social programmes.

Evaluation research embodies a political dilemma – that the price to be paid for insisting on methodological rigour on the grounds both of scientific quality and democratic accountability may well be a lack of utility for policy-makers (Hope, 2002a). The *Martinson Problem* (arising from an earlier era of evaluation of sentencing and penal treatment), is that the rigour of evaluation research is perceived by policy-makers to heighten the probability of reaching ‘Nothing Works’ conclusions. What may serve the public interest in one way – i.e. by holding government accountable – may hinder it in another – i.e. by inhibiting government from finding out “what works”, thereby inhibiting political progress. There would thus be considerable advantage to be gained if it were possible to preserve the safeguards of validity inherent in ‘scientific method’ while at the same time producing useful results of “what works”. Nevertheless, when push comes to shove, the greater resources of persuasion available to politics than to science can mean an uneven contest in which methodology comes to be sacrificed to political expediency (Hope, 2004).

For all its promises and inducements to social scientists, the ideology of EBPP can be used as a means of exerting political control over the conduct of evaluation research in order to validate and legitimise government policy (Hope, in press). EBPP invokes the aura of science to gain support and consent for policy (Campbell, 1969). In this sense, the ‘evidence’ generated from evaluation research, resting as it does on apparently scientific foundations, can be used politically as an unassailable means of supporting Government policies (Hope, 2006). As experienced during an evaluation

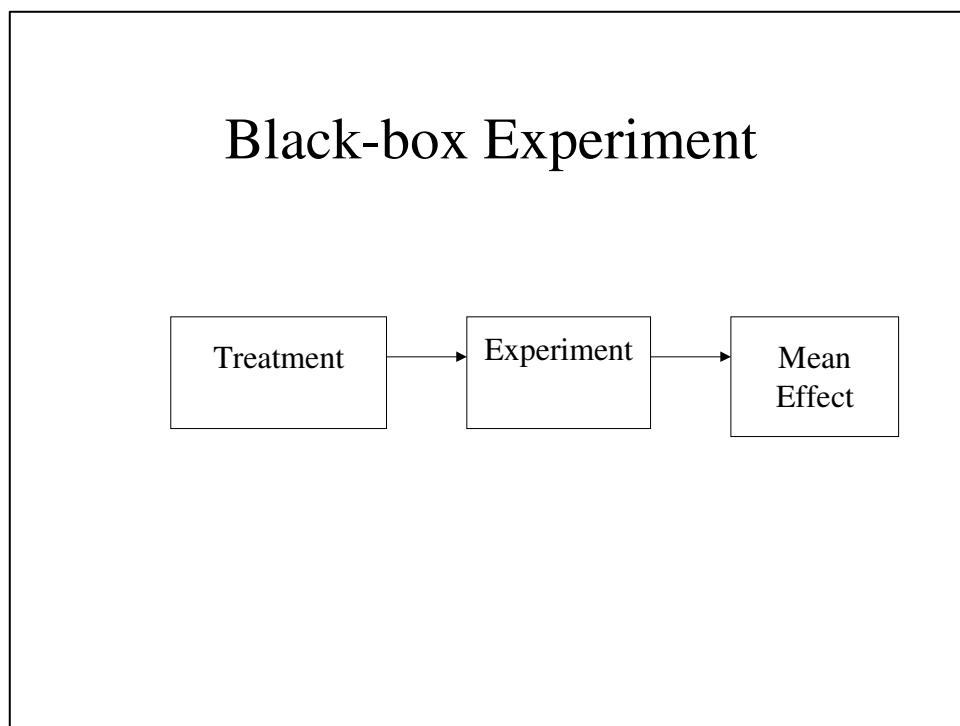
of part of the Reducing Burglary Initiative of the Home Office Crime Reduction Programme (1999-2003), this stratagem puts pressure on evaluation researchers to produce the ‘correct’ findings (STC, 2006), leading contracted researchers to compromise on research quality standards; and government to commission alternative congenial research (Hope, 2004) and/or to ignore, distort, or suppress the findings of the research that it has commissioned (Hope, 2008).

The logic of evaluation research

The purpose of evaluation research is causal attribution – specifically, to assess the extent to which the deliberate implementation of policy intervention X changes substantially an identified crime/security phenomenon Y. Attaching a *value* to this assessment (cost-effectiveness) is an important, though secondary, purpose since it depends upon the identification of the scale of effect. The methodological task is to identify the unique effect of the intervention, net of other influences. Thus, evaluation methodology has the two-fold task of estimating the unique effect of an intervention, while ensuring that such estimates are valid representations of true effects, primarily by protecting against bias in the production of the estimates.

The simplest conception of an evaluation is in the form of a ‘*black box*’ *experiment* (see Figure 2), where a particular intervention is conceptualised as a ‘treatment’ applied to a group of subjects and an estimate is obtained of the mean effect of the treatment on the subject group. The treatment is causally prior to the effect.

Figure 2



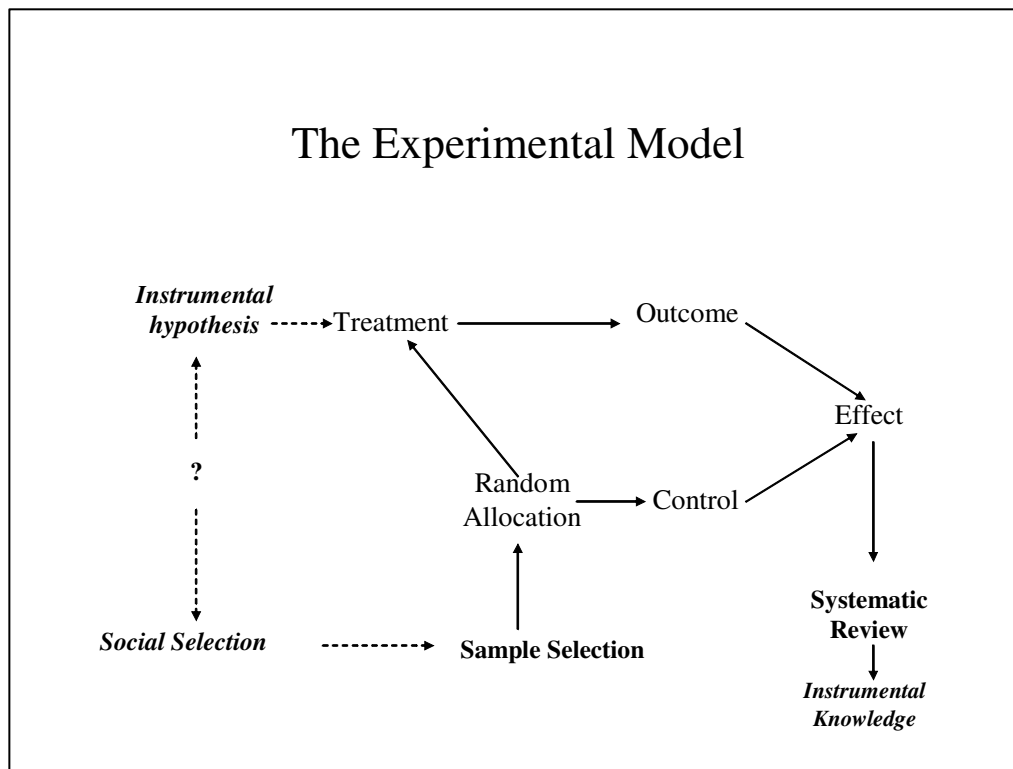
Evaluation methodology has evolved to solve the *fundamental evaluation problem* that this conception poses. As Heckman and Smith (1995) put it, this arises from:

“...the impossibility of observing what would happen to a given person [or community] in both the state where he or she receives a treatment (or participates in a program) and the state where he or she does not. If a person could be observed in both states, the impact of the treatment could be calculated by comparing his or her outcomes in the two states, and the evaluation problem would be solved”. (*ibid*: 87)

Evaluation methodology should thus be concerned with the operationalisation of methods that support the valid estimation of the effect of being subject to an intervention, net of alternative, *counter-factual* conditions that the subject could experience. Empirically, we only have available the observable ‘fact’ of being subject to the intervention. Since we cannot observe what would have happened if the intervention had not happened (i.e. the counter-factual condition), we have no certain means of knowing what factors ‘caused’ the specific conditions observed in the subject during and after the intervention. Because the counter-factual is unobservable, we cannot evaluate the actual probability of the effect occurring against the unactualised probability of it not occurring.

In the experimental design (Figure 3), those persons assigned to non-exposed conditions represent the counterfactual condition because the process of random assignment to the ‘exposed’ and ‘non-exposed’ conditions means that members of either group are ‘exchangeable’ and that the subsequently observed differences between the two represents the difference between the factual and the counterfactual, and hence indicates causation (Hernán and Robins, 2006). It should be noted that ‘random allocation’ to the control condition does not imply that the experience of the control condition is itself ‘random’. Unlike in a laboratory, evaluation research applied to social policy interventions takes place in ‘field-settings’ where both the experiences of the subjects and the intervention are fundamentally embedded in social life, hence the term *quasi-experimentation* (Cook and Campbell ,1979). Random allocation, therefore, cannot create a vacuum by disembedding subjects from their social context; rather, in principle, it ensures that subjects are exchangeable (i.e. have identical experiences) so that observed differences can only be due to the unique experience of the intervention, which is thus an *abstracted* aspect of experience that only makes sense in terms of the logic of the methodology used to abstract it.

Figure 3



Yet if randomisation is crucial to the capacity to control the estimation of effects, then it is also critical to be able to control randomisation. Community crime prevention interventions differ from criminal justice interventions in terms of the styles of *governance* associated with the different institutional settings and programmes typically implemented within each sphere (Hope, 2005c). In practice, this shapes the ability to exercise control in order to bring about subjects' compliance with the intervention, and hence defines the nature of the intervention itself. Community interventions tend to be premised upon mechanisms of *collective, voluntary participation* - for example, where a community group, or non-state agency acting on its behalf, is concerned with collective self-advancement; or where a government is concerned with the physical, economic or social 'regeneration' of a residential area. In contrast, the 'treatment' programmes taking place in courts and corrections (including supervision in community settings) are characterised by *individual, constrained participation*. Typically, they are concerned with altering the rewards and/or sanctions bearing upon individuals so that they can 'fit-in' to prevailing institutional arrangements. Here, ideas about intervention arise from different spheres - of criminal justice, education and 'correction' - that are, at root, the responsibility of the State - the institution with legitimate, judicial powers of coercion over subjects.

Subjects' participation in community programmes is voluntary because these involve the social and institutional arrangements of civil society. The mode of governance in civil society is indirect and diffuse - holders of power (government) have to govern 'at a distance', influencing and shaping the setting itself, utilising a range of essentially non-coercive measures, in order to engineer the compliance of private citizens, including influencing the future behaviour of 'putative' as distinct from

actual offenders or victims⁴. In contrast, settings such as courts, prisons and (only to a lesser extent) schools, and even labour markets (particularly of the unemployed or low waged), include subjects whose participation is largely *involuntary* or over whom power-holders (essentially the state and its agencies) have some considerable (ultimately coercive) power of *control* and influence⁵. These latter settings differ therefore from civil society communities in that they comprise the various arenas where there is a greater level of resource for engineering the compliance of subjects available to agencies with the authority and legitimacy to bring about change.

The capacity to *control* subjects directly is crucial not just because it shapes the policy conditions that define the intervention but also because it affects the *research conditions* that will define the most appropriate evaluation research methodology. Two conditions of control are crucial: the ability to (a) allocate subjects to specific research conditions, and (b) command the compliance of agents in implementing policy interventions (treatments). The criteria employed to assess methodological quality in the Scientific Methods Scale (SMS, noted above) rely principally upon the power of control. For instance, a systematic review based upon the SMS quality standards, under the heading “what doesn’t work?” listed “...community mobilization of residents’ efforts against crime in high-crime, inner city areas of concentrated poverty” (Sherman *et al.*, 1998, p. 8), citing Hope (1995) as corroboration (erroneously) of this proposition. Yet because of the differential availability of coercive control in community relative to criminal justice settings, the application of the SMS review methodology does not tell us to what extent this conclusion is due to the problem of operationalising SMS-favoured methods (of randomisation and control) rather than the relative efficacy of each sphere to intervene in crime. Worse, comparatively unfavourable inferences could be drawn simply on the basis of comparing flawed efforts at experimentation in community settings with competently-designed and well-executed experiments in settings more amenable to the manipulation and control of human subjects (Hope, 2005c).

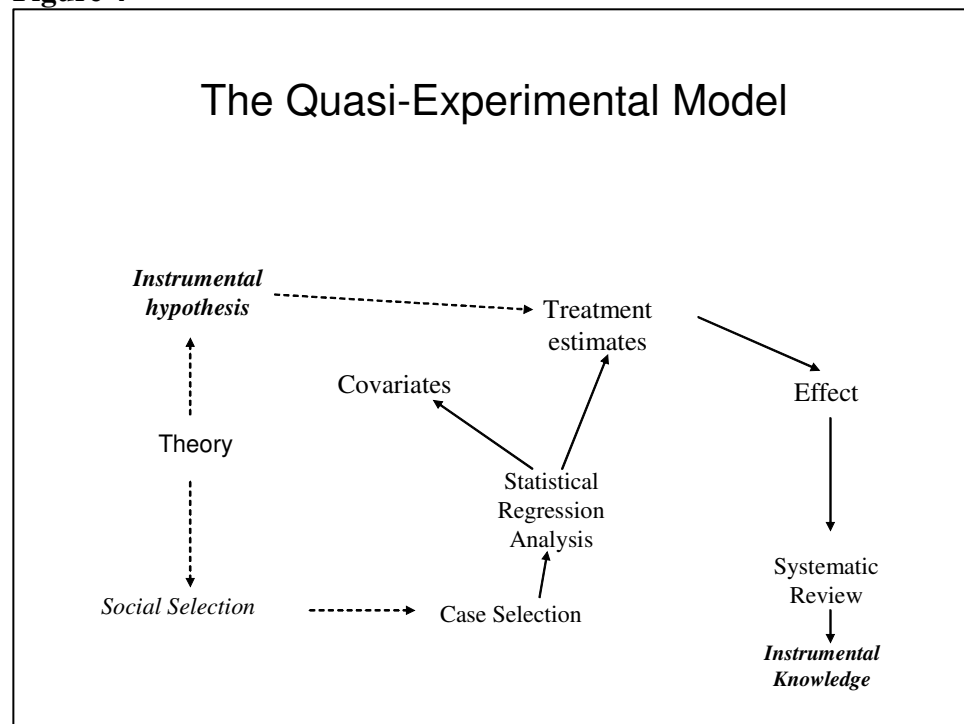
Methodology of evaluation research

Quasi-experimental methods have evolved to compensate for the difficulties of randomisation-by-design encountered by applying experimental designs in field-settings (Figure 4). Essentially, they rely upon statistical adjustment to compensate for the absence of the capacity to control by design (Judd and Kenny, 1981). The evaluations of Neighbourhood Watch (Study 2), Community Policing (Study 3), Priority Estates Project (Study 4), DICE (Study 7) and Community Policing (Study 9) have all employed a similar design: non-randomised matching of ‘experimental’ and ‘control’ sites, using regression adjustment methods applied social survey data to estimate treatment effects, net of the effect of measured covariates (i.e. social characteristics of residents correlated with the outcome variable) that might differ between experimental and control sites (Farrington, 1997).

⁴ That is, they accord to the sphere of ‘primary crime prevention’ “identifying conditions of the physical and social environment that provide opportunities for or precipitate criminal acts (Brantingham and Faust, 1976; p. 290).

⁵ Similarly, these accord to the sphere of ‘secondary crime prevention’ (Brantingham and Faust, 1976).

Figure 4



The methodological contribution of Campbell and his followers lay in advocating research design solutions, short of the pure experiment, to the many various ‘threats’ to the validity of programme evaluation research (Cook and Campbell, 1979). Essentially, these threats stem from the practical difficulties of implementing valid experiments with human subjects in field-settings. In this tradition there are four classes of threat to validity (Farrington, 2003):

1. *Statistical conclusion validity*: whether the hypothesis of a causal relationship between intervention and outcome is supported by inferential statistical tests. As in all statistical hypothesis testing, there are two test criteria: testing the probability that the intervention hypothesis should be accepted (Type I Error testing); and testing the probability that the null hypothesis should be rejected (Type II Error). Information to assist in these decisions rests, respectively, upon the calculation of statistical significance (via the appropriate statistical test) and of the size of effect (statistical power calculation) (Farrington, 2003).
2. *Internal validity*: the key question of inference of a causal relationship between intervention and outcome.
3. *Construct validity*: “...the adequacy of the operational definition and measurement of the theoretical constructs that underlie the intervention and the outcome” (Farrington, 2003: 54).
4. *External validity*: the degree and extent to which we can have confidence that the results from a particular intervention can be generalised to other contexts.

While there has been much debate about the way in which evaluation methods can address these issues, these classes of validity still remain a clear categorisation of the nature of the methodological problems encountered in evaluation research.

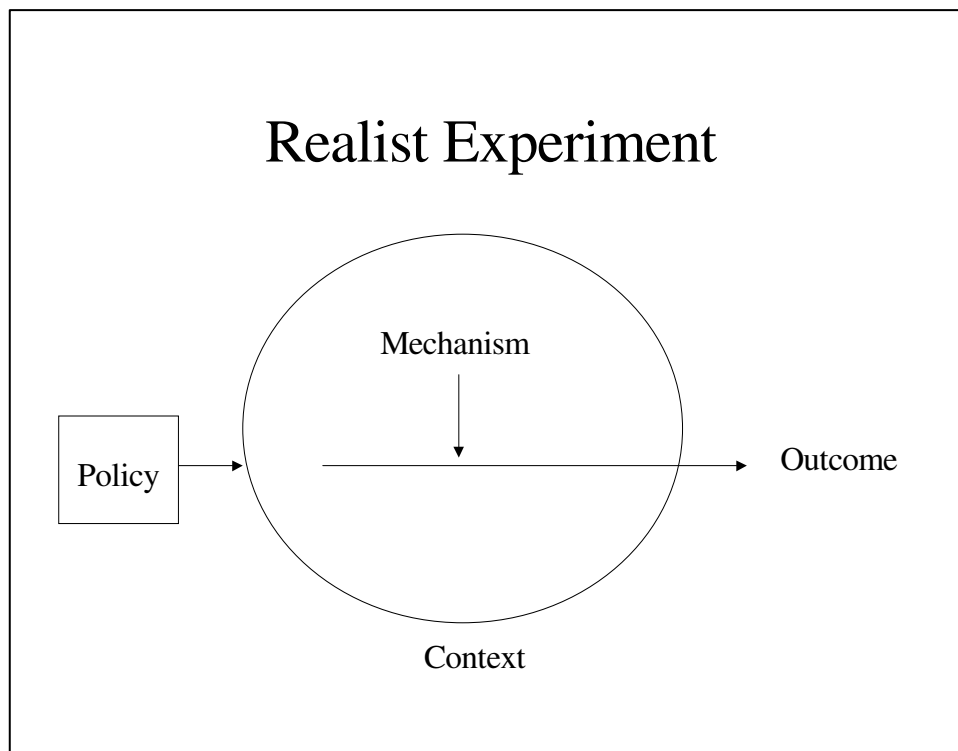
The realist critique

A crucial difference between a community crime prevention initiative and, say, a controlled laboratory experiment is that, in the former, the ‘treatment’ has far less integrity *a priori*. In field-settings, any ‘treatment’ intervention is *produced* over time in specific contexts, in complex practical and organisational ways, utilising varying combinations of authority, capital and resources. Moreover, any intervention always acts upon a specific context (such as a residential community) and produces its effect often through a variety of *mechanisms* intrinsic to the intervention. Study 3 (Bennett, 1991) was criticised by Pawson and Tilley (1994) on the basis that its ‘black-box’ model of experimental intervention may have concealed a variety of mechanisms within the intervention that might have an effect on the outcome; while the statistical discounting of covariate differences between sites conceals the likelihood of different contexts having separate effects on outcomes (see also Bennett, 1996; Pawson and Tilley, 1996). Essentially, their critique is concerned with establishing the greater importance of Construct and External Validity, which they suggest has greater political and practical relevance, than social scientists’ primary concerns with Internal and Conclusion Validity.

Nevertheless, there are incommensurate and contradictory trade-offs amongst the four types of validity (Judd and Kenny, 1981). Thus, the critique of the Random Controlled Trial (RCT) applied to field-settings is that although it maximises Internal and Conclusion Validity, it risks both greater artificiality and oppressive/unethical practice (usually both together). Further, in being a practical intervention, in order to implement an RCT, it is necessary to abstract it artificially from the social and institutional contexts in which its subjects are embedded. Thus, the dilemma of the RCT is that the more it maximises internal validity the more it loses on external and conclusion validity; the dilemma for the realist critique is the opposite.

Pawson and Tilley (1997) propose an alternative way of conceptualising the production of an outcome by hypothesising that any outcome is always the product of a particular *context-mechanism-outcome* (CMO) configuration (Figure 5). The task of evaluation is therefore to uncover the variety of CMOs inherent in an intervention and then seek additional similar cases (replication) where regularities of policy input can be discerned through comparison between variations in CMOs – essentially by comparing sequences and collections of case-studies (e.g. Tilley and Webb, 1994). Tilley (1993), for example, reports two case-studies that purport to ‘replicate’ the Kirkholt Burglary Reduction Project, as part of a process to assess the effectiveness of a policy of repeat burglary prevention (Pawson and Tilley, 1997). Further replications are also held to vindicate the Kirkholt thesis (Farrell, 2005).

Figure 5



Nevertheless, this so-called scientific realist method threatens the Internal and Conclusion validity of its findings. In particular, it encourages (or insufficiently guards against) *bias* in its selection of case studies and in its investigation of causal mechanisms. These difficulties stem both from its methodology – that is, a reliance on qualitative, case-study methods - and its epistemology - that is, its inductive method of enquiry. For example, ‘replication’ evidence from samples of CMOs (Hope, 2002a), analysis of national trends (Hope, 2007) and evaluation studies (Hope *et al.*, 2004) - have found no further instances to date of the mechanism thought to be central to the apparent success of the Kirkholt Project – that a change in either the prevalence of repeat victims, or the frequency rate of (repeat) victimisation, has any significant association with a change in the overall crime (incidence) rate. If anything, it would seem that the crime prevalence rate (the proportion of victims in the population) has a far greater impact on the crime incidence rate than does the frequency (repetition) rate of victimisation (Hope, 2007; 2002a).

The ‘realist’ interpretation of the success of the Kirkholt Project is *ex post facto*. This opens up ‘realist’ evaluation to the charge of selective interpretation of evidence (a fallacy of the inductive method of reasoning) which is not helped by the absence of research design and comparison in the case-study method commonly employed. One issue concerns that of *intent* – without establishing what was intended prior to implementation, it would not be possible, logically, to establish that what had happened was what had been intended. There is always the danger of the *ex post facto* fallacy (*post hoc, ergo propter hoc*), which is particularly worrying in policy evaluation: if original intentions are not met, it may be politically attractive to take credit retrospectively for what had actually happened fortuitously, irrespective of any

particular intention (Campbell, 1969). These problems are compounded by two well-known problems of selectivity (Hope, 2006; Hope et al., 2004):

1. *Selection bias*: the difficulty of ‘control’ in community crime prevention (noted above) means that projects to be evaluated are never selected at random – more usually, they have been selected to participate on some basis that may be related to the outcome. Candidates for evaluation may then have already been ‘cherry-picked’ for inclusion on their likelihood or promise of success (e.g. if crime has been going down previously, or if the implementing agents seem enthusiastic, co-operative, competent and efficacious). The problem is compounded, first, when projects are selected on a competitive basis; and second, when they are given a ‘pre-launch improvement’ – both of which occurred during the Reducing Burglary Initiative (Study 8) (Hope, 2004)⁶. A further consequence of EBPP is that it justifies the reporting only of the selection of ‘successful’ case studies, biasing the result obtained and inhibiting learning (which may need to analyse failure as much as success).
2. *Regression to the Mean*: national crime prevention programmes may be targeted on high crime areas, on the basis that a greater preventive gain might be attained relative to expenditure, as was the case with many projects in the Crime Reduction Programme (Study 8). That said, it is then difficult to distinguish the extent to which any subsequent reduction is due merely to the phenomenon of regression to the mean, at least by the methods usually employed (Hope, 2006; Campbell, 1969).

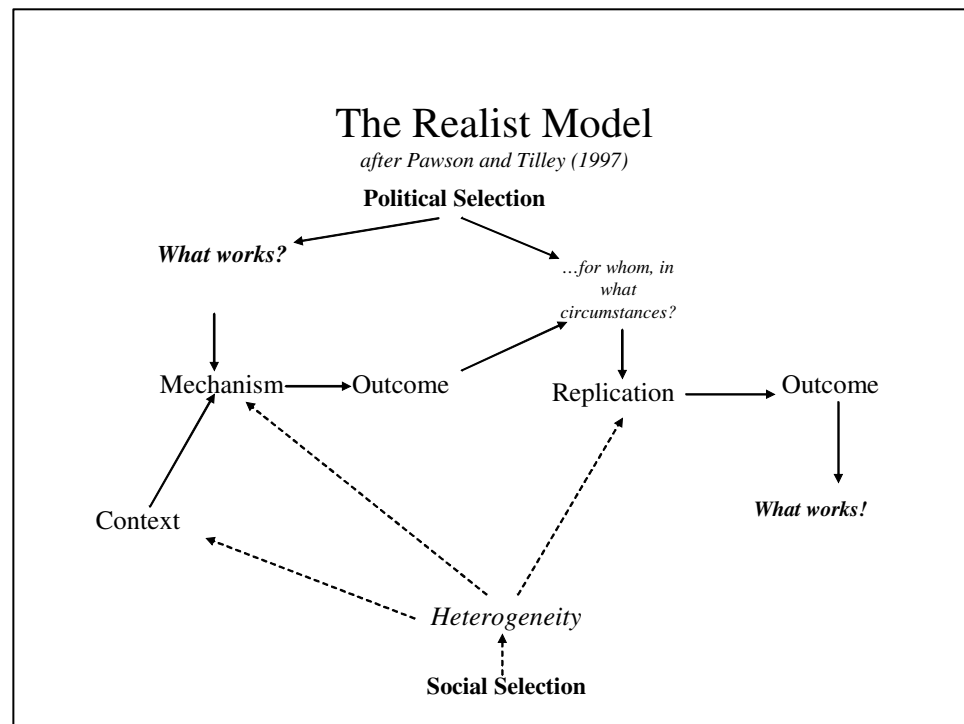
A related logical problem is that of *expectation*. The customary antonym in crime prevention policy to “what works” is “nothing works” – that is, the expectation that action will *either* have a positive benefit (crime reduction) *or* a nil-effect. What this antonym cannot incorporate is the possibility of unintended but still nevertheless significant *negative* or harmful consequences as a result of crime preventive intervention. To the extent that those who propose crime-reductive actions are oriented towards political utility, the logical possibility of unintended consequences - that actions might cause harm or are more costly than the problem they sought to remedy - cannot be countenanced without undermining their rationale of crime prevention. This mode of thought does not allow for the logical possibility of crime *permissio*n: that the specific actions of the programme, regardless of whether or not they are intended to reduce it, may nevertheless permit or encourage crime to increase. Nevertheless, evidence is available from the Reducing Burglary Initiative (Study 8) that the particular activities of a particular crime prevention project – ironically focussing specifically on repeat burglary prevention – appeared to have brought about a 38 per cent *increase* in the rate of burglary victimisation, as well as in the rate of repeat victimisation (Hope, et al., 2004)⁷.

In sum, the realist model fails to account for both political selectivity and social selectivity and thus becomes vulnerable (whether intentionally or not) to bias (Figure 6).

⁶ And since, as it turned out, neither effort actually guaranteed success, sponsors were simply prompted to re-analyse the data so as to come up with sought-after outcomes (Hope, 2004).

⁷ Another occasion for retrospective re-analysis (see fn. 6)!

Figure 6



Programme Mechanisms

Despite its vulnerability to bias in selection and interpretation, the realist approach, with its emphasis on the production of prevention, does point to the importance of mechanisms – or rather causal sequences – developing over time to generate crime prevention outcomes, which the typical ‘black-box’ experimental approach fails to capture. Theories of the generation of community crime prevention, based on ideas about collective efficacy and social capital, posit causal sequences whereby contextual variables interact with intra-community social processes to produce outcomes developing over time (Hope and Karstedt, 2003; Hope, 1995). Such processes were identified by the ‘mixed methods’ approach of the Priority Estates Project Evaluation Study (Study 4) (Foster and Hope, 1993). Social capital theories posit a set of necessary *interaction effects* in order to produce community safety; but it is precisely these kinds of interaction effect that the assumptions of the SMS quality criteria would construe as ‘threats to validity’⁸. Broadly, these take the form of two kinds of so-called threat:

1. *Selection x Treatment*. This could itself take two forms: first, where the intervention programme cannot be separated from the social and political structures in which target communities are embedded. For example, only certain kinds of community are likely to support ‘neighbourhood watch’ community organisations, thus state-agencies interested in establishing them will select only those kinds of communities that provide fertile soil for neighbourhood watch groups to flourish, while social processes within such

⁸ In the evaluation research literature these are often referred to as ‘moderator’ and ‘mediator’ effects on treatment (Farrington, 2003; 1998), though in individual-treatment interventions these appear not to be as severe a set of ‘threats’ as they do in community crime prevention.

communities will also be ‘working’ to capture state resources, creating ‘club goods’ of security for themselves (Hope and Trickett, 2004). Arguably, the nil effect identified in evaluation studies of Neighbourhood Watch (Study 2 above) may be due to the similar levels of indigenous social capital present in the matched experimental and control communities. Second, as a consequence, a kind of ‘elective affinity’ (to use Weber’s term) is established between treatment and selection, so that the one cannot operate without the other, to the extent that the treatment itself comes to be defined in terms of the selection criteria (i.e. the characteristics of the setting).

2. *Treatment x Outcome*. The objective of the SMS quality scale is to ensure that it is the treatment that causes the outcome – termed *Conclusion I* (Judd and Kenny, 1981). However, there are two other relations between treatment and outcome that, although they more accurately define the process of change, through a sequence of mediating processes, paradoxically reduce the correlation between treatment and outcome that SMS designs endeavour to maximise, for instance, *Conclusion II* where the treatment causes the potential mediator. Here, the treatment intervention into a community may actually seek to enhance the ‘mediating variable’ – for example, the development of the right mix of social capital. Thus, the intervention creates opportunities for bridging capital to be developed, say, by creating intra-community organisations, utilising and empowering community leaders (already possessing ‘linking’ social capital), which then empowers the community by creating additional linking capital for community members, that enhances their efforts to create bridging institutions, and so on (Hope and Karstedt, 2003). There is also *Conclusion III* – that the sum of these activities (e.g. collective efficacy) itself becomes the mediating variable, such that when it is controlled statistically, or in research design, the direct correlation between the treatment variable (the originating governance intervention) and the outcome variable (i.e. crime) disappears, even though it has actually had an important, though indirect, causal influence.

Evidence of both these ‘threats’ to explanation is apparent from Study 4 (Hope, 2005c). Thus, the paradox is that were the evaluation research design to control for these apparent threats to validity – or that an SMS review rejected the evidence from studies that did not seem to have these qualities – it would be likely to reject the influence of the treatment on the outcome when, in practice, it had a very significant effect. In this case, it would be the evaluation method itself, and not the intervention, that produced the result, thus leading to Type II error inferences – that the treatment had no effect, when in fact it did.

A related issue in community crime prevention is that of *spill-over* – whether displacement of crime or diffusion of benefit. Space precludes a detailed discussion of measurement issues involved, though again, it is necessary to consider methodological issues concerning the validity of observations of apparent spill-over effects. For instance, Hope et al. (2004 – Study 8) found, paradoxically, that more projects achieved a greater net reduction in crime in their wider area, including a surrounding ‘buffer zone’ than in their specific target area. Yet this was more likely to be a spurious rather than a true diffusion of benefit since projects were actually part of bigger crime reduction initiatives covering wider areas beyond the target, and that

they were likely to affect the behaviour of populations of offenders, victims and community members in the communities beyond their relatively small target areas.

Programme implementation

The realist critique also points to the importance of intervention process in the estimation of causal effect. The process of implementation is not only crucial to a project but also a logical pre-requisite to evaluation – whatever the rigour of evaluation methodology, it will not be possible to detect the effects of an intervention if little or nothing has been implemented to produce those effects (Hope, 1985). Generally, outcomes are likely to be affected by the:

1. *Efficacy of crime prevention measures* (this is the traditional concern of programme developers);
2. *Tractability of the target/community context*: i.e. how easy or difficult it is to implement the crime prevention measures, given the characteristics of different environmental and community contexts; and
3. *Efficiency of the organisation of delivery*: i.e. how efficient or productive are the agents and their procedures in implementing the intervention.

Each of these independent factors is likely, separately and in combination, to interact to influence the probability of a specific outcome of an intervention (Hope *et al.*, 2004). Rather than being separate issues, the evaluation of implementation is necessary condition for the evaluation of outcome. Yet it is also necessary to separate out its effects in order to ascertain programme outcomes, or more usually failures, adducible to failures of implementation (Hope, 2005b).

Implementation issues have arisen in most British community crime prevention programmes (Hope and Murphy, 1983). This has led to two methodological responses:

1. *Process studies of implementation*: evaluation studies of the CRP Reducing Burglary Initiative (Study 8) were obliged to undertake detailed process studies, moving-on from mere description and monitoring in order to explain specific implementation outcomes, provide inputs to cost-effectiveness studies (Bowles and Pradipto, 2004) and generate a set of conditions that might account for differential impacts. For instance, the relative failure of projects to reduce burglary was related to a differential willingness to abandon and adapt initial intervention plans which, in turn, was related to the organisation of the project and the characteristics of the lead agency. In particular, projects that were dominated by the police seemed notably inflexible, persisting with ineffectual implementation plans that eventually lead to failure (Hope *et al.*, 2004).
2. *Measurement of project intensity of action*: following pioneering work by Ekblom *et al.* (1996, Study 6), evaluation studies in the RBI (Study 8) developed various quantitative measures of project intensity of action which could measure the amount, quantity, duration and tempo of implementation.

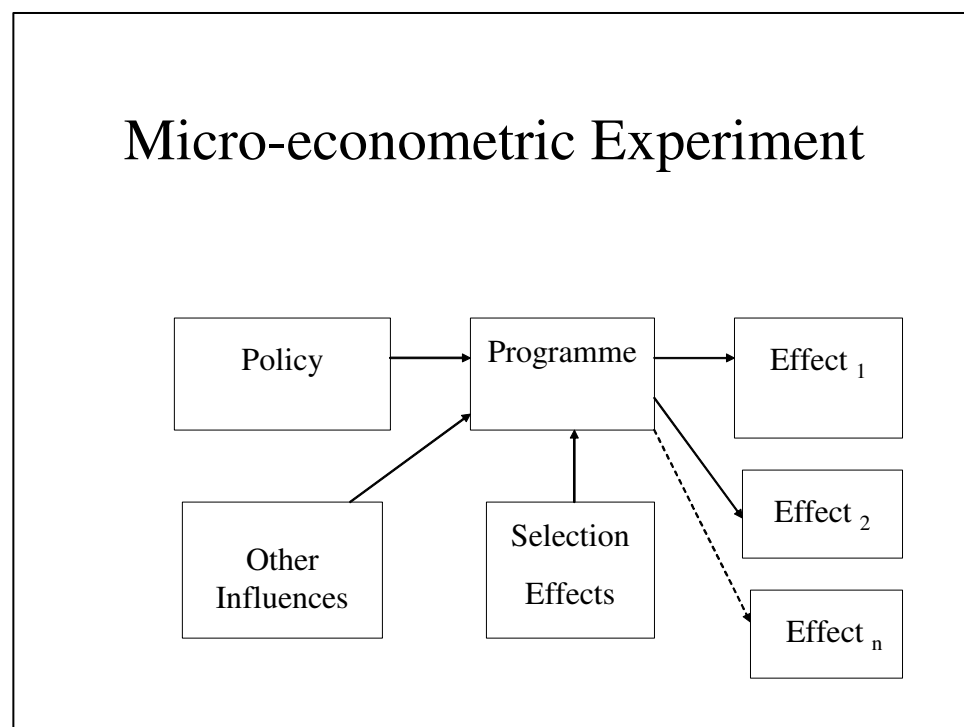
Hope *et al.* (2004, Study 8) combined both of the above approaches to create an Intensity of Action score – consisting of a periodic, quantitative, cumulative measure of the impact of implementation during and beyond the duration of the project.

Observational data on implementation, administrative and cost data were combined into a Calendar of Action from which a quantitative measure could be derived, which was immensely useful in econometric, time-series analysis of programme impact (Crawley and Hope, 2003).

Conclusion: heterogeneity and selection bias

The realist critique both illustrates and succumbs to two methodological problems of non-experimental research that have a major effect on the inferences to be drawn from evaluation research: *heterogeneity* and *selection bias*. Contexts, interventions and outcomes are heterogeneous (as the realists argue) but non-experimental evaluation research also needs to build-into its design and analysis protection against inferential error due to manifold selection bias (of which realists are wholly guilty). As both interventions and evaluations are embedded in the social world, quasi-experimental research design is insufficiently powerful on its own to isolate generalisable estimates of programme effect free from selection bias. Building on the work of Heckman (2001) in micro-econometrics, evaluation research needs to model selection effects, isolate unique programme effects, and identify heterogeneous outcomes (Figure 7).

Figure 7

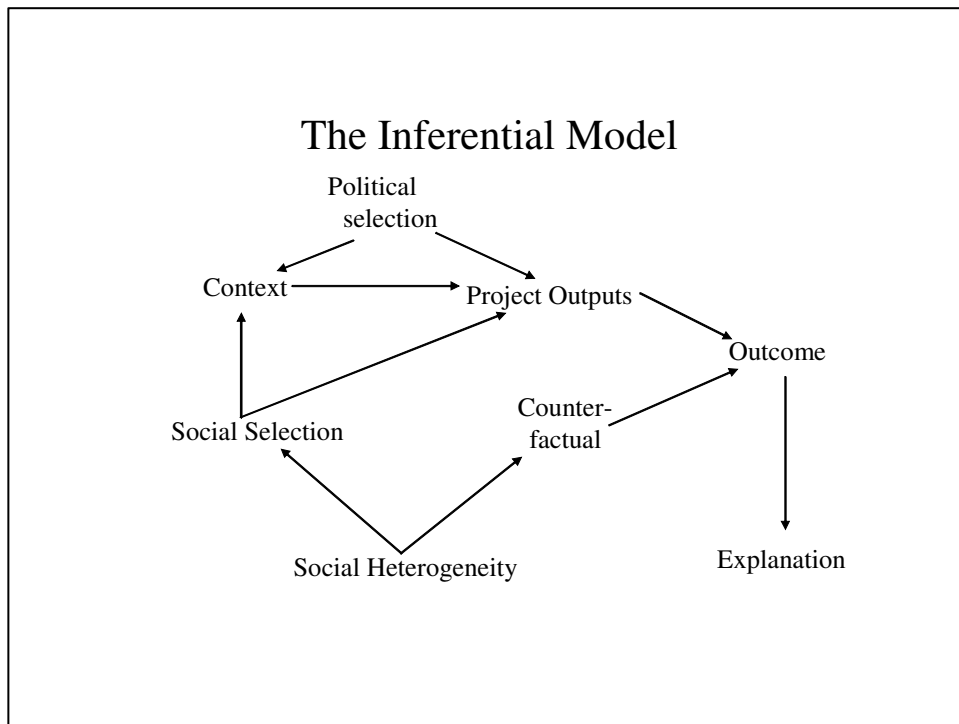


Hope et al. (2004) for instance, fitted separate multivariate time-series models to the crime trends in each project area in their study, with appropriate specifications for the particular properties of the time series in each project (capturing heterogeneous local patterning of crime), and then estimated unique project effects (measured by the Intensity of Project Action variable), net of *ex ante* projections of the crime trend due

to other influences. Ekblom et al. (1996) carried out a highly-sophisticated study to adjust observed data to account for a large range of selection biases and heterogeneous effects that could not be controlled by research design.

In conclusion, what may be needed to advance understanding of the impact of community crime prevention are *inferential statistical models* that seek to estimate selection bias and heterogeneity – properties fundamentally affecting evaluability – using counter-factual causal reasoning to underpin micro-econometric analysis (Figure 8).

Figure 8



References

- Bennett, T. (1991). 'The effectiveness of a police-initiated fear-reducing strategy'. British Journal of Criminology, 31, 1-14.
- Bennett, T. (1990), *Evaluating Neighbourhood Watch*. Aldershot: Gower.
- Bowles, R. and R. Pradipto (2004). Reducing Burglary Initiative: an analysis of costs, benefits and cost effectiveness. Home Office Online Report 43/04
- Campbell, D.T. (1969) 'Reforms as Experiments', *American Psychologist* 24: 409–29. (Reprinted as Campbell, D.T. (1978) 'Reforms as Experiments', in J. Bynner and K.M. Stribley (eds) *Social Research: Principles and Procedures*. Oxford: Oxford University Press).
- Connell, J. P. and A. C. Kubisch (1998). 'Applying a theory of change approach to the evaluation of comprehensive community initiatives: progress, prospects and problems' . In K. Fulbright-Anderson, A.C. Kubisch and J. P. Connell (1998). New Approaches to Evaluating Community Initiative. Volume 2: Theory, Measurement, and Analysis. Washington, DC: The Aspen Institute.
- Cook, T.D. and D.T. Campbell (1979) *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin.
- Crawley, P. and T. Hope (2003). 'Measuring the outputs of crime reduction projects'. Community Safety Journal, 2 (3), 4-12.
- Ekblom, P., H. Law and M. Sutton (with P. Crisp and R. Wiggins) (1996). Safer Cities and Domestic burglary, Home Office Research Study 164. London: Home Office Research and Statistics Directorate.
- Farrell, G. (2005). 'Progress and prospects in the prevention of repeat victimisation'. In N. Tilley (Ed) Handbook of Crime Prevention and Community Safety. Cullompton, Devon: Willan.
- Farrington, D. P. (2003). 'Methodological quality standards for evaluation research'. The Annals of the American Academy of Political and Social Science, 587: 49-68.
- Farrington, D. P. (1998) 'Evaluating 'Communities That Care': Realistic Scientific Considerations', Evaluation, 4: pp. 204 - 210.
- Farrington, D. P. (1997) 'Evaluating a Community Crime Prevention Program'. Evaluation, 3: pp. 157 - 173.
- Forrester, D., M. Chatterton, and K. Pease (1988) The Kirkholt Burglary Prevention Project, Rochdale. Crime Prevention Unit Paper 13. London: Home Office.
- Forrester, D., S. Frenz, M. O'Connell, K. Pease (1990). The Kirkholt Burglary Prevention Project: Phase II. Crime Prevention Unit Paper 23. London: Home Office.

Foster, J. and T. Hope (1993). Housing, Community and Crime: the impact of the Priority Estates Project. Home Office Research Study No. 131. London: HMSO.

Gill, M. and A. Spriggs (2005). Assessing the impact of CCTV. Home Office Research Study 292, London: Home Office Research, Development and Statistics Directorate.

Gladstone, F.J. (1980) <http://www.homeoffice.gov.uk/rds/pdfs05/hors62.pdf>

Goldblatt, P. and C. Lewis (Eds.) Reducing Offending: an assessment of research evidence on ways of dealing with offending behaviour. Home Office Research Study 187. London: Home Office. Pp. 51-62.

Heckman, J. (2001). 'Microdata, heterogeneity, and the evaluation of public policy: Nobel Lecture. Journal of Political Economy, 109, 673-748.

Heckman, J. and J.A. Smith (1995). 'Assessing the case for social experiments'. The Journal of Economic Perspectives, 9, 85-110.

Hernán, M.A. and J.M. Robins (2006). 'Estimating causal effects from epidemiological data'. J. Epidemiol. Community Health, 60, 578-586.

Hollin, C. R. (2008). 'Offending behaviour programmes: Does only randomization glister?' Criminology and Criminal Justice, 8: 89 - 106.

Hope, T. (In press). 'The first casualty: evidence and governance in a war against crime'. In P. Carlen (Ed) Imaginary Penalties. Cullompton, Devon: Willan.

Hope, T. (2008). 'A firing squad to shoot the messenger: Home Office peer review of research'. In W. MacMahon (Ed.). Critical Thinking about the Uses of Research. Evidence Based Policy Series 1. London: Centre for Crime and Justice Studies, King's College London. www.crimeandjustice.org.

Hope, T. (2007). 'The Distribution of Household Property Crime Victimization: Insights from the British Crime Survey' In M. Hough and M. Maxfield (Eds.), Surveying Crime in the 21st Century. Crime Prevention Studies Vol. 22, Monsey, NY: Criminal Justice Press and Collumpton, Devon: Willan

Hope, T. (2006). 'Things can only get better'. Criminal Justice Matters. No. 62 Winter 2005/06 <http://www.kcl.ac.uk/depsta/rel/ccjs/cjm62.html> . (Reprinted in House of Commons Science and Technology Committee Scientific Advice, Risk and Evidence Based Policy Making, Seventh Report of Session 2005-06, Volume II Oral and Written Evidence, Ev 145. The House of Commons: HC 900-II 8 November 2006).

Hope, T. (2005a). 'The new local governance of community safety in England and Wales'. Canadian Journal of Criminology and Criminal Justice, 47 (2), 367-387

Hope, T. (2005b). 'Sustainability in crime prevention: a nautical tale.'. 10th Anniversary Colloquium. International Centre for the Prevention of Crime, Paris, 30 November – 2 December, 2004. International Observer, Edition 36, March 2005.

Montreal: International Centre for the Prevention of Crime. http://www.crime-prevention-intl.org/io_view.php?io_id=117

Hope, T. (2005c). 'Pretend it doesn't work: the 'anti-social' bias in the Maryland Scientific Methods Scale'. European Journal on Criminal Policy and Research, 11, 275-296.

Hope, T. (2004). 'Pretend it works: evidence and governance in the evaluation of the Reducing Burglary Initiative'. Criminology and Criminal Justice, 4 (3), 287-308.

Hope, T. (2002a). 'The road taken: evaluation, replication and crime reduction'. In G. Hughes, E. McLaughlin and J. Muncie (Eds.) Crime Prevention and Community Safety. London: Sage Publications. (ISBN 0-7619-7409-1), pp. 37-57.

Hope, T. (2002b). 'La riduzione della criminalità, la sicurezza locale e la nuova filosofia del management pubblico'. Dei Delitti e della Pene, Quadrimestrale Anno IX, n. 1-2-3 (seconda serie) Gennaio-Dicembre, 2002, pp. 207-229

Hope, T. (1998). 'Community crime prevention'. In C. Nuttall, P. Goldblatt and C. Lewis (Eds.) Reducing Offending: an assessment of research evidence on ways of dealing with offending behaviour. Home Office Research Study 187. London: Home Office. Pp. 51-62.

Hope, T. (1995). 'Community crime prevention'. In M. Tonry and D.P. Farrington (Eds.) Building a Safer Society: strategic approaches to crime prevention. Chicago: University of Chicago Press. Crime and Justice: a review of research, 19, 21-89.

Hope, T. (1985). Implementing Crime Prevention Measures. Home Office Research Study No. 86. London: HMSO. <http://www.homeoffice.gov.uk/rds/hors1985.html>

Hope, T. and Susanne Karstedt (2003) 'Towards a new social crime prevention'. In H. Kury and J. Obergfell-Fuchs (Eds.) Crime Prevention: New Approaches. Mainz, De.: Weisse Ring Verlag-GmbH. pp. 461-489.
<http://www.pbni.org.uk/bsthopeskarstedt.pdf>

Hope, T. and D.J.I. Murphy (1983) 'Problems of implementing crime prevention: the experience of a demonstration project'. The Howard Journal, 22, pp 38-50.

Hope, T. and R. Sparks (2000). 'For a sociological theory of situations (or how useful is pragmatic criminology?)'. In A. von Hirsch, D. Garland and A. Wakefield (Ed.) Situational Crime Prevention: ethics and social context. Oxford: Hart Publishing.

Hope, T. and A. Trickett (2004) 'Angst Essen Seele Auf...but it keeps away the burglars! private security, neighbourhood watch and the social reaction to crime'. Kölner Zeitschrift für Soziologie und Sozialpsychologie, Sonderheft 43, 441-468.

Hope, T., J. Bryan, E. Crawley, P. Crawley, N. Russell and A. Trickett (2004). Strategic Development Projects in the Yorkshire and the Humber, East Midlands and Eastern Regions. Home Office Online Report 41/04. London: Home Office.
<http://www.homeoffice.gov.uk/rds/onlinepubs1.html>

Hirschfield, A. (2004). The impact of the Reducing Burglary Initiative in the North of England. Home Office Online Report 40/04.

Judd, C.M. and D. A. Kenny (1981). Estimating the Effect of Social Interventions. Cambridge: Cambridge University Press.

Kodz, J., K. Pease and M. Stephens (2004). Explanatory note for online reports OLR 40/04 to OLR 43/04. Home Office Online Report 39/04

Knox, J., P. Pemberton (PA Consulting Group), and P. Wiles (2000). Partnerships in Community Safety. London: department of the Environment, Transport and the Regions.

Maguire, M. (2004). 'The Crime reduction programme in England and Wales: reflections on the vision and reality'. Criminology and Criminal Justice, 4, 213-237.

Millie, A. and M. Hough (2004). Assessing the impact of the Reducing Burglary Initiative in southern England and Wales. Home Office Online Report 42/04

O'Malley, P. (2000). Risk, crime and prudentialism revisited. In K. Stenson and R. Sullivan (Eds.) Crime, Risk and Justice: the politics of crime control in neo-liberal societies. Cullompton, Devon: Willan.

Pawson, R. and N. Tilley (1998a). 'Caring Communities, Paradigm Polemics, Design Debates', Evaluation, 4: pp. 73 - 90.

Pawson, R. and N. Tilley (1998b). 'Cook-Book Methods and Disastrous Recipes: A Rejoinder to Farrington', Evaluation, 4: pp. 211 - 213.

Pawson, R. and N. Tilley (1997). Realistic Evaluation. London: Sage Publications.

Quinton, P. and J. Morris (2008). Neighbourhood policing: the impact of piloting and early national implementation. Home Office Online Report 01/08

Sherman, L.W., D.C. Gottfredson, D.L. MacKenzie, J. Eck, P. Reuter and S. Bushway (1997). Preventing Crime: What Works, What Doesn't, What's Promising. Washington, D.C. US Office of Justice Programs.

STC (2006). Oral Evidence (Professor Tim Hope), Wednesday 24 May 2006. House of Commons Science and Technology Committee Scientific Advice, Risk and Evidence Based Policy Making, Seventh Report of Session 2005-06, Volume II Oral and Written Evidence, Ev 38. The House of Commons: HC 900-II 8 November 2006.

Tilley, N. (1993). After Kirkholt - Theory, Method and Results Of Replication Evaluations?, Crime Prevention Unit Series Paper No. 47. London: Home Office.

Tilley, N. and G. Laycock (2000). 'Joining up research, policy and practice about crime'. Policy Studies, 21, 213-227.

Tilley, N. and Webb, J. (1994). Burglary Reduction: Findings From Safer Cities Schemes, Crime Prevention Unit Series: Paper No.51. London: Home Office

Tuffin, R., J.Morris and A.Poole (2006). An evaluation of the impact of the National Reassurance Policing Programme. Home Office Research Study 296. Home Office Research, Development and Statistics Directorate.