A Framework for Deriving Policy Implications from Research

Dr Aidan Wilcox
Professor Alex Hirschfield
September 2007

CSB/14
Queensgate
Huddersfield
Hd1 3DH
01484 473676

Applied **Criminology** Centre

The University of Huddersfield

	CONTENTS	PAGE

A FRAMEWORK FOR DERIVING POLICY IMPLICATIONS FROM RESEARCH

Abstract. Scopic (social contexts of pathways into crime) is a five year ESRC funded programme of research which aims to achieve a better understanding of how young people become involved in crime. A key consideration in a research programme of this type is how best to generate policy implications from academic research. This report sets out a proposal as to how this might be achieved. It begins with a brief description of scopic; we then set out a framework for deriving policy implications which takes into account the validity of the research, and the practicality, feasibility and ethics of proposed policy options. Consideration is then given to how policy implications might be made concrete. The report concludes with a case study in which the model is applied to one of the early published research papers from scopic and we comment on the strengths and weaknesses of this approach.

1   INTRODUCTION

Governments, like individuals, increasingly turn to science as a basis for the decisions or policies they make. Although the policy making process is arguably far from rational, and faces competing influences, including financial and legal constraints, ideology and values, rigorous evidence forms an important part of the policy making process, at least if the rhetoric of 'evidence-based' policy is to be believed.[1] There is, for example, a long history of the use of empirical research to aid the development of health policy. In criminal justice, by contrast, the relationship between evidence and policy has been more tenuous and cyclical, and there have been many examples of policies being influenced more by public opinion, ideology or pressure groups than by evidence (Wilcox, 2005; Wilcox and Hirschfield, 2007).

In recent years, however, government has expressed a commitment to be 'guided by the evidence' (Blunkett, 2000) and this has come to be known as the 'what works' approach in criminal justice. The increasing publicly-expressed commitment to evidence provides the context for this report and raises the question as to how best can the results of empirical research be fed into the policy process. In an attempt to answer that question, we outline a framework for deriving policy implications. Our aim here is to raise some of the questions which need to be answered, rather than to provide the definitive answers themselves.

The first stage of the process involves an assessment of the rigour or validity of the research findings, in particular the strength of causality demonstrated, the generalisability of the results and the extent to which the variables used were accurate measures of the theoretical constructs. Secondly, the causal variables are then assessed as to whether they are amenable to change through policy intervention, and if so, whether it would be feasible to do so. Thirdly, the ethicality of changing the variables is considered, before concrete proposals for policy change are suggested. Before describing the model in more detail, we begin by introducing the scopic studies.

---

[1] An approach to policy making in which initiatives are to be supported by research evidence and policies introduced on a trial basis are to be evaluated in as rigorous a way as possible (Plewis 2000: 244).

1.1 The scopic studies

Scopic is an ESRC funded, five year, multi-site study of the pathways into and out of crime. The network comprises four British universities and a number of international, collaborating universities. The main aim of scopic is to achieve a better understanding of the factors associated with onset and desistance from offending through research conducted with three different age groups. It is hoped that this will unravel the relationship between individual characteristics and behavioural and social contexts and assist in the development of public policies to reduce criminality and enhance individuals' life chances.[2]

The four UK universities are focused on discrete but complementary aspects of the research, and it is worth briefly describing their roles.[3]

The Institute of Psychiatry at *Kings College London*, is investigating the role of environmental risk (E-risk) in the origins of anti-social behaviours amongst children under the age of 10. The research, led by Professor Moffitt, employs a twin study design, following up 1116 sets of identical twins selected to be representative of twin births during 1994/5. The 'E-risk' twin study is funded by the Medical Research Council, and the ESRC provided additional funding to allow KCL to undertake a survey of neighbours of the twins in the study. Data from the neighbourhood survey will be used to investigate whether neighbourhood quality is correlated with children's problem behaviours because it has a causal influence on child development or because neighbourhoods reflect the genetically-influenced risk characteristics of families who live in them.

The Institute of Criminology, at the *University of Cambridge*, under the direction of Professor Wikström, leads the consortium and its research involves a longitudinal study targeting the developmental period (ages 14-15) where offending peaks. The Peterborough Adolescent Development Study (PADS) follows a sample of 707 12-year-old boys and girls from the period of low involvement in crime to peak involvement and uses follow up interviews and survey data. The research tests the hypothesis that social mechanisms influencing age-related offending vary with community context, and that differences in individuals' routines, processes of decision-making and perceptions of alternatives play a significant role in accounting for those variations. As with the KCL study, PADS itself follows on from a previous cross sectional study of 2000 14-15 year olds in Peterborough, known as the Peterborough Youth Study.

The third empirical component of scopic is the *University of Sheffield's* pathways out of crime project (SPOOCS) led by Professor Bottoms, and is a study of desistance from crime. Focusing on persistent offenders in their twenties from Sheffield, the research adopts a 'life-course' approach. Repeated interviews with offenders are used to explore how the interaction between individual characteristics (such as impulsivity) and aspects of current lifestyle influence individuals' decisions to desist from crime. The research builds on previous research at Sheffield University on the study of crime in its urban context.

---

[2] Descriptions of the projects are drawn from the official website. For further information about scopic see: www.scopic.ac.uk

[3] There are, in addition, four international universities engaged in collaboration with scopic. These are the Chicago Neighbourhoods project, the Pittsburgh Youth Study, the Montreal Two Samples Longitudinal Study and the Zurich Project on the Social Development of Children (see http://www.scopic.ac.uk/studies.htm).

The role of the *University of Huddersfield* in relation to Scopic is twofold. The first, which is the topic of this report, is to identify how policy implications of research such as that being undertaken by SCOPIC members could be derived. The second is to study the inter-relationship between policy, research and government with a view to making recommendations designed to enhance the influence of research on policy. The second aim is pursued in our report '*Pathways into Policy: A Study of the Relationship Between Research, Policy and Government'* (Wilcox and Hirschfield, 2007.

2   POLICY, ANALYSIS AND EVALUATION: SOME CONCEPTUAL ISSUES

Before setting out our framework for deriving policy implications, it is worth considering some of the analyses that can be distinguished that are relevant to the formulation, implementation and understanding of policy.

The first of these is *'analysis for policy';* that is to say research that sheds light on the nature of an activity (e.g. youth offending) and the factors that facilitate or inhibit its occurrence. The analysis of offending patterns and the isolation of processes and mechanisms through which they can potentially be explained, help to identify the possible approaches to its management and reduction. This type of 'problem diagnosis' can be used to point policy in the right direction. The object of the analysis is primarily the problem to be solved, although in some cases, existing policies might also be part of the problem. Codd (1988) distinguishes between two approaches to analysis for policy. The first, 'policy advocacy', involves the researcher making specific policy recommendations; while the second, 'information for policy', requires researchers to provide policy makers with information to enable them to revise or formulate policy (Codd, 1988: 235).

'Policy analysis' (or as Gordon et al, 1977 call it, 'analysis *of* policy') is entirely different. This 'policy analysis' is concerned with making the policy process more rational through the description of policy problems, clarification of goals, identification of policy options, and evaluation of alternatives (Haas and Springer 1998; Colebatch 2005). Here, the object of the analysis is the policy (or policies) that potentially influence the problem under study. Thus policy analysis is concerned with attributes of policy interventions that have relevance to a condition and how these inter-relate. A number of policies may be involved: they may be funded by different government departments with different levels of funding; their timing may overlap; there may be similarities in their aims, tactics or target groups. Alternatively, they may differ in some respects (e.g. in their timing or intended beneficiaries) or bear very little relationship to one another, they may be complementary, they may duplicate the work of other policies or they may contradict each other.

Thus 'policy analysis' should be about documenting and understanding the aims, objectives, tactics and timing, together with the spatial and socio-demographic targeting, of *all* relevant policy interventions potentially influencing a common outcome (e.g. reductions in youth offending). In practice, this form of analysis is rare, mainly because of the absence of consistently recorded data on policy interventions.

Whilst a considerable amount of attention is focused on the measurement, analysis and mapping of need (e.g. deprivation, ill health and crime) virtually no attention is given to analysing the dispensation of policy at the local or community-level. Thus the extent to which there is synergy, duplication or contradiction between different policy approaches to youth offending at the community level is largely unknown. Combining 'analysis for policy' with 'policy analysis' should enable the researcher to begin to identify some important questions, such as:

- Are there any mechanisms/processes that research has identified which none of the relevant policies are tackling?

- Are there any mechanisms/ processes that policies are prioritising that research shows are not problematic?

5

- Where is there a need for innovative policy responses that are more closely aligned to the nature of the problems to be solved?

Policy evaluation is different again, and fulfils an 'instrumental' function by answering the question how far a policy has achieved its aims and objectives, for whom, how and at what cost (Sanderson, 2000:438). The focus can be on a single policy intervention or on a group of interventions that constitute a policy initiative or programme. Commonly, evaluation studies are conducted as ex post evaluations (i.e. at or towards the end of a policy initiative) rather than concurrently or in advance of implementation.

Typically, policies that are evaluated emanate from a single government department, agency or partnership. The primary focus is on the performance of a designated initiative. Although the question of how far other policies are confounding factors in evaluation studies is often posed, the data and techniques for unravelling the complexity of policy dispensation and its effect on attribution of significant change (e.g. in offending) are lacking. In short, evaluation studies are often hampered by the failure to conduct a comprehensive policy analysis.

These different forms of analysis are usually conducted in research silos. None of them individually can be considered to be a sufficient policy analysis model. Evaluation studies can be too policy-centred and fail to see the complexity of the problem. Pure research can be too problem-focused and fail to see the importance and complexity of policy making. It is only through cross-referencing and integrating aspects of these analyses that a comprehensive policy analysis model can begin to be developed.

Two further distinctive areas of activity can be distinguished, namely, policy design and appraisal and policy delivery (implementation).

Policy design is concerned with identifying and planning appropriate responses (e.g. to offending) based on theory, anticipated mechanisms, intended beneficiaries and desired outcomes. It is about identifying a relevant intervention or package of measures to impact upon a problem. The emphasis is on the content of the intervention; what action to take and which tactics to use; when, how, where, for whom and by whom. In practice, the interventions need to be the right course of action given the nature of the problem and the context within which it occurs. Policy design should also be informed by findings from evaluation studies that identify the effectiveness of different interventions. Part of the process may involve some form of appraisal or feasibility exercise aimed at scrutinising the interventions for consistency, value for money and likely effectiveness but policy design in primarily about putting together the components of the policy rather than problem diagnosis or policy evaluation.

Effective policy needs to be well designed but also implemented properly. *Policy delivery* is concerned with tactics, targeting, project and resource management, inter-agency working as well as data sharing and other aspects of implementation. Policies may fail to impact upon offending because of theory failure (doing the wrong thing given the nature of the problem and/or the context) because of implementation failure (attempting to do the right thing, but badly) or because of both (doing the wrong thing inefficiently). There are abundant examples of policies that have failed, not because the theories and ideas that underpinned them were flawed, but on account of implementation failure (e.g. Gill et al, 2006). Thus the components of delivery, for example, project planning, resource and staff management, adherence to timetables, have to be sound.

In the following table we summarise the different types of analysis described above.

Table 1 Policy and types of analysis

Term	Description
Analysis for policy	Research into the nature and manifestation of a phenomenon / problem (e.g. offending) including exploration of possible causal mechanisms in order to inform ameliorative policies
Policy analysis	An examination of how different policies likely to influence a given outcome (e.g. offending) relate to each other in terms of their aims, tactics, resources, timing, location and intended beneficiaries.
Policy design	The formulation of a policy (i.e. objectives and tactics) for tackling an identified problem with some initial assessment of the feasibility of implementation and / or likely cost effectiveness
Policy delivery	All aspects of policy implementation including forward planning, management, data sharing, expenditure, timetable, targeting of resources, performance monitoring
Policy evaluation	An assessment of the impact and effectiveness of a policy in achieving stated outcomes. This may be undertaken concurrently with policy implementation of retrospectively (ex post evaluation)

Our aim in this report is to think about how policy implications could be derived from research. Of the various types of analysis described above, the model we propose fits most closely into the category of *analysis for policy*. That is to say, the findings of empirical research can help to inform the definition of the problem and identification of possible solutions. In the sections which follow we set outline the main steps involved in deriving policy implications from research, starting with a consideration of the validity of research.

3 A MODEL FOR POLICY MAKERS

This chapter outlines the four main strands to our model for deriving policy implications from research. The first considers issues of methodological quality, as these relate to internal, external and construct validity. The second section considers the feasibility of changes to policy implied by research, while the third looks at how ethical issues might be resolved. Finally we discuss how policy implications can be made concrete.

3.1 Is the research methodologically sound?

Drawing out the policy implications from research is inseparable from establishing its trustworthiness, or validity (Edwards, 2000). If one accepts that not all research is of equal quality, and that only evidence of the highest quality should be fed into the policy process, then it is clearly important that policy makers be able to classify and rate the methodological rigour of such studies, in order to assess what weight to place on the reported findings. Given the growing number of empirical studies in criminology, it has become increasingly necessary for policy makers and academics to be able to classify and rate the methodological rigour of such studies, in order to assess what weight to place on the reported findings. As one of government's chief social science research advisors commented recently 'not all research is of sufficient quality to inform policy' (Davies 2004: 3). The aim of this section is to sketch out how one might assess the overall quality of a study, and key to this is the issue of validity.

Validity concerns the extent to which a given proposition, inference or conclusion approximates to the truth (Trochim, 2000). In other words, when researchers make a claim based on their data, for example that poor educational achievement leads to involvement in crime, the validity of the claim relates to how likely this is to be true. There are a number of different types of validity, but the most relevant for our purposes are internal, external and construct validity.[4]

3.1.1 Internal validity
Internal validity addresses the truth of the question as to whether a change in one variable (e.g. drug use) has led to a change in another (e.g. reconviction) and is of relevance to any study which attempts to establish a causal relationship. An internally valid piece of research is therefore one which has demonstrated a cause-effect relationship between variables, by a process of ruling out alternative plausible explanations (Wilcox, 2005). For evaluative research (such as reconviction studies) the rules of internal validity are relatively straightforward. There are numerous examples in various areas of scientific research of rating scales or checklists, which attempt to distil and measure the essential elements of a rigorous research design.

Rating scales were initially developed to assess the quality of randomised controlled trials (RCTs) in medicine. One of the first such scales was developed by Chalmers and colleagues (1981) for medical studies. They argued that in order to assess the quality of a RCT, the publication would need to contain details about: the nature of the intervention; the methodology of the study; how data were analysed; as well as presenting results in such a way so as to enable comparison with the results from other studies (Chalmers et al 1981). For them, the most important determinants of internal validity were the blinding procedures. Ideally, both patients and doctors

---

[4] There is also descriptive validity (the adequacy of the reporting of the methodological and other key aspects of a study) (Farrington 2002) and statistical conclusion validity (the truth or otherwise of the statistical influences made) (Trochim 2000).

should be unaware of who has been allocated to which condition, and participating doctors should also be blinded with regard to ongoing trial results (Chalmers et al 1981).

Since this first scale was developed, numerous others have been designed for medical studies (according to Moher et al, 1995, by 1993 a further 24 scales had been developed, in addition to nine 'checklists'). Most, if not all, of the medical scales developed up to that time had been concerned only with the rating of randomised experiments. However, in medicine (and even more so in social science) there are many evaluations of a quasi-experimental nature. Downs and Black (1998) therefore set out to explore the feasibility of designing a scale to measure quality of methodology in both randomised and quasi-experimental studies. Although quasi-experimental methodology differs from that of the randomised controlled trial, they both have in common an intervention, possible confounding variables and an outcome of interest. The aim of both is to test for a causal relationship between the intervention and outcome and to minimise the threats to internal validity (Downs and Black 1998).[5] The scale they developed included 26 items, and unlike many of the scales developed for RCTs, Downs and Black included items relating to external validity, statistical conclusion validity, as well as the quality of reporting.

Rating *scales* developed later in the social sciences. However, Logan, writing in 1972, proposed a checklist of seven criteria which he considered to be the minimum requirements for evaluative research of penal treatments. He included the establishment of similar treatment and control groups, adequate definition of variables and the importance of replicability. However, it was not until 1989 that the first rating scale specifically for the social sciences was developed. Gibbs' scale (1989) for social work studies identified 14 quality components, and assigned points to each to produce an overall score ranging from zero to one hundred. Most of the criteria he identified related to the methodological rigour of the study (such as the use of randomization, rate of attrition and the size of treatment groups) although a fifth of the points were awarded for aspects of the reporting of the study (i.e. descriptive validity), such as the quality of the description of the intervention, its aims and target group.

Brounstein et al (1997) used a different approach in developing a rating scale for alcohol and drug abuse prevention programmes. They rated studies on a scale of zero to five on each of 10 criteria, including adequacy of sampling and sample size, adequacy of comparison groups, adequacy of statistical analyses and testing of alternative explanations.

Perhaps the most influential rating scale to date for criminological evaluations is the Maryland Scientific Methods Scale (SMS), developed by Sherman and colleagues (1998). They drew upon previous rating scales for the social sciences, in particular Brounstein et al's scale. The scale was used to rate studies as part of a major investigation into 'what works' in (US) criminal justice interventions. The scale used many of the same quality components as in Brounstein et al's scale, and was scored in a similar way. Each of the quality items was scored from one to five and the summary score (from one to five) was based on the assessor's overall interpretation of study quality. Finally, the first author developed a scale as part of his doctoral research (Wilcox, 2005) and this rated studies on the following factors: the comparability achieved between groups; rates of attrition; statistical analysis, and; quality of implementation.

---

[5] Bias is the tendency for results to depart systematically from the 'true' result. Unbiased results are said to have internal validity (NHS CRD 2001: 2.5.2).

Not all research which is of potential value to policy makers is of the experimental or quasi-experimental nature which such rating scales have tended to focus upon. Indeed, many of the questions of policy interest are not amenable to exploration through experimental or other quantitative approaches. Some of the research conducted for Scopic, for example, is not of this nature, being based on longitudinal surveys or interview data. The issue of how to measure internal validity is thus more problematic, and the assessment of validity becomes somewhat more subjective. That is not to say one cannot assess internal validity, indeed there are certain types of question one can pose in relation to most research designs, including the following:

- Recruitment – is there anything in the selection of cases which is likely to introduce systematic bias to the results?
- Loss of cases due to drop out – is the rate of attrition such that the results are likely to be unrepresentative?
- Statistical analysis - has the analysis been conducted in such a way as to minimise bias? One might consider here issues such as sample size and adequacy of reporting of the statistical tests for example.

The assessment of qualitative research is particularly open to debate. However, a number of authors have attempted to systematise the assessment of methodological quality in qualitative research. Guba and Lincoln (1989), for example, propose five dimensions along which the reliability of answers derived from qualitative research can be judged. These are credibility (whether the findings give a true account of what respondents said), transferability (how applicable the findings are to other settings), dependability (how well documented are the methods and decisions taken by the researchers), confirmability (whether the findings are backed up by evidence) and authenticity (whether the research increases understanding of the issue).

The Cabinet Office, in conjunction with the National Centre for Social Research has developed a comprehensive framework for assessing the credibility, rigour, and relevance of individual qualitative research studies (Spencer et al, 2003). Their framework sets out a number of 'appraisal questions' structured around the broad headings of research findings, design, sample, data collection, analysis, reporting, reflexivity and ethics. These appraisal questions include:

- How credible are the findings
- How well does the evaluation address its original aims and purposes?
- How clear is the basis of evaluative appraisal?
- How defensible is the research design?
- How well was the sample and data collected?
- How well has the detail, depth and complexity of the data been conveyed?
- How clear are the links between data, interpretation and conclusions?[6]

One subtype of evaluation is the 'theory-based' evaluation. This approach to evaluation involves breaking down the theoretical assumptions on which an intervention is based into a sequence of causes and effects, and collecting data to examine how well each step is borne out. There are parallels with the realist evaluation methodology described by Pawson and Tilley (1997) and both share the advantage of being able to produce results during the lifetime of the project, and by

---

[6] See further:
http://www.policyhub.gov.uk/evaluating_policy/quality_framework/framework_table.asp

making explicit the assumptions embedded in a programme, they also help strengthen the evaluation's influence on policy makers. Granger discusses theory based evaluations, and argues that their validity rests on whether the evaluators have described and discussed credible counterfactuals at all stages of their work (Granger, 1998). To achieve this, the evaluation must include strong theories, representative samples, multiple methods and designs, serious testing of alternative hypotheses, and plans for replication (Rosenbaum 2002).

Stevenson and Cooper (1997) argue that the main determinant of the quality of research, whether qualitative or quantitative, is the extent to which the researcher reflects on their own research. This reflexivity allows researchers to choose a methodology appropriate to the research question of interest and allows others to critique their choice. Internal validity is not the only concern when determining the utility of research findings. Its wider applicability (external validity) is also of central concern.

3.1.2 External validity

External validity concerns the extent to which the causal relationship in a study (if established) can be generalised to other places, people, times, settings or operational definitions of the variable – in other words to whom do the findings of the study apply? (Farrington, 2002). External validity is important because policy makers constantly have to make decisions as to whether to introduce policies based on research conducted in other times, countries or settings. If the findings are relevant only to a very specific population, then the policy implications may be of limited use.

Unlike internal validity, which can be established within one study, the strength of external validity is increased by the number of trials of the intervention conducted. There are two main reasons why repeated tests of effectiveness are needed in order to demonstrate external validity. The first concerns the risk of drawing erroneous conclusions from one study due to the possibility of a 'false positive' result. Given the generally accepted levels of statistical significance ($p<0.05$), one would expect that around one in twenty trials of an ineffective policy would show a positive result by chance alone. If one had two trials of an intervention showing the same result, the risk of a false positive falls to one in 400. The second reason is that there is a trade-off between certainty of causation (internal validity) and of generalisability (external validity). The more closely controlled and rigorous the experiment, the greater the danger that the results do not hold with respect to the uncontrolled world in which policies are implemented Tenbensel (2004). In other words, the results of any one trial may be due to factors specific to that site, sample or time period, and thus not generalisable to other settings. The more representative the sample in a particular study is of the population in question, then, the more confident one can be that the results will be generalisable to that population (Cook and Campbell, 1979). However, the main test of the external validity of any research findings is repeated testing, of different operational definitions, within different settings and on different populations (Sherman and Gottfredson, 1998).

Whilst there is much research into effectiveness in the area of situational crime prevention, replication remains difficult, with contradictory results commonplace. It is the intrinsic variability of the social world that limits the external validity of much research (Tilley and Laycock, 2000). A famous example is the Minneapolis domestic violence study, which seemed to suggest that mandatory arrest of offenders was more effective than non-arrest (Sherman and Berk, 1984). The results of this one study were quickly taken up by authorities in many US states who enacted mandatory arrest legislation. However, when the experiment was subsequently replicated over six other US sites, these failed to reproduce the original findings

(Binder and Meeker, 1992). Policy makers were left none the wiser, therefore, as to what the effects would be of implementing a mandatory arrest policy.

Similar excitement was generated following an evaluation of cognitive skills programmes for offenders, which seemed to indicate that the programme led to a significant reduction in reconviction rates (Friendship et al 2002). These programmes were subsequently rolled out across the prison service. Again, however, a replication of the evaluation failed to find a 'treatment effect' (Cann et al 2003).

It has been suggested that one reason for such findings is that replications of successful programmes have attempted too literal a translation, while ignoring that subtle changes in the new setting may render the intervention ineffective. Ekblom (2002: 144) and Pawson and Tilley (1994) argue that successful replication requires the identification of which causal mechanisms interact with which aspects of the context to produce the desired crime preventive outcome and the utilisation of this knowledge in the design and implementation of the replication. Of course one problem with adopting such an approach is that any one crime prevention initiative may engage multiple mechanisms, and it becomes impossible to conduct enough evaluations to explore the full range of interactions between them. This is what Cullen and Gendreau describe as the problem of 'unending specification'; in other words arguing that rehabilitation would work only if programmes could be developed which address the 'unending permutation of offender-treatment-type-setting permutations' (Cullen and Gendreau, 2000: 129).

One further issue for policy makers is how to utilise the findings of specific studies of mechanisms and of more general theory. While the former may produce more readily understandable and usable knowledge, they will have narrow applicability. On the other hand, the implications of theories utilising more general mechanisms and contexts may have applicability to a wider range of crime problems, but be more difficult to operationalise (Ekblom, 2002).

It was clear from our interviews[7] with policy makers that they recognised the problem of using research findings from other places, times and settings. As one commented:

> Evidence which is not necessarily UK, it's really difficult to know what to do with international evidence, because you know, what's, take America, America's quite a different country from this, but is more similar than Japan, so but it's very difficult to know, ok this is Sweden, lots of stuff coming out of the Netherlands, Scandinavia, quite different, but are they are different as Japan, I don't know, I find it very difficult to judge on the international landscape what it, everything is useful in terms of ideas, but what is useful in terms of potentially transferable practice is very difficult. (HO VN)

Assessing the external validity of research is probably the most subjective of the areas considered here, as a judgement has to made as to how similar the circumstances of the original research are to the current context. That is not a question that can simply be answered by the application of a rule or rating scale.

3.1.3 Construct validity
In the social sciences the definitions and measurements of constructs are subject to debate. Construct validity refers to the extent to which the operational definitions of

---

[7] For details of interviews conducted as part of our scopic research, please see the report 'Pathways into Policy: A Study of the Relationship between Research, Policy and Government'.

cause and effect accurately reflect their theoretical constructs (Farrington 2002). Criminal justice interventions are often assessed by their impact on offending behaviour, yet this is not possible to measure directly (absent 24 hour surveillance and knowledge of offenders' intent). Thus one is reliant on official records, self report, or, in the case of children under the age of criminal responsibility a proxy for offending, such as anti-social behaviour. The further the measurement used is removed from the actual behaviour, the more questionable is the utility of the variable employed. Similarly, one needs to consider whether the measurement of explanatory variables under investigation (e.g. educational attainment) closely represent their theoretical constructs. Evidently, the ability to judge the construct validity of the variables will depend partially on the adequacy of the description of the variables in the report.

3.2   Are the implied policy implications feasible?

Coleman (1975) draws the distinction between 'policy variables', that is those variables which can be or have been subject to policy control, and 'situational variables', in other words variables that play a part in the causal relationship with the outcome of interest, but which are not open to policy control (Coleman, 1975: 24).

There are many examples of both policy variables and situational variables in the crime prevention literature. Age and gender, for example, are both strong predictors of an individual's risk of reconviction, but neither of these could be changed by means of policy. These 'static' risk factors are therefore situational variables. 'Dynamic' risk factors, such as employment status, drug use or accommodation are also associated with reconviction (May, 1999). These latter are clearly policy variables, as it is possible to change an individual's employment status, for example, by means of public policy. It is important to consider, therefore, whether the constructs measured in empirical studies reflect policy or situational variables.

Once one has identified that a variable is open to change, the question becomes is it feasible or practical to do so? In answering this question, one would have to consider issues such as the likely cost of the policy, and the expected size of the benefit if it was successful. Ideally, of course, one would want to conduct a cost benefit analysis. However, for research which is more exploratory, which is not concerned with testing policies, but with seeking to uncover causal mechanisms, then the question of practicality is therefore somewhat subjective in nature and necessarily provisional.

It may be that the causal factor has only been uncovered due to the in-depth nature of the research, and would not normally be expected to be revealed otherwise. For example, in depth interviewing of mothers and reports from teachers may indicate that maternal depression causes antisocial behaviour in their children. One has to ask how practical it would be to screen all mothers for depression and then successfully to treat it in order to ameliorate their children's behaviour. Similarly, many aspects of parenting style (e.g. level of parental supervision) which are associated with antisocial behaviour in children (e.g. Farrington 1992) would also be very difficult to measure and then to change by means of public policy. It is recognised that in some instances it may not be possible to provide an answer to the question of the feasibility of a policy option without further research, for example by piloting an initiative.

3.3   Are the implied policy changes ethical?

An important consideration is whether a policy, even if possible and practical, is desirable or ethical. Whether consciously or not, those involved in the policy making process must deal with the issue of ethics. There are few, if any, crime prevention policies which can be implemented in a moral vacuum (Shaftoe, 2004). Policy makers may see their role as identifying a moral ideal, to which policies should aspire, or a morally acceptable minimum that, realistically, could be adopted as policy, or even a political compromise that can arguably meet ethical analysis. The answer is likely to vary between individuals and differences of opinion will be manifested in the debates over particular policy recommendations.

The findings of research are not neutral; indeed they raise important ethical issues, particularly in the area of crime prevention. Although it may not be immediately evident, the efficacy of a proposed policy is an ethical issue. This is because funding for the criminal justice system (as in all areas of public expenditure) is limited. By taking the decision to fund one policy (for example mentoring programmes for young offenders) policy makers, by implication, decide not to fund other possible programmes (e.g. after-school clubs). If the former policy turns out to be less effective, or more harmful, than the latter, then this represents a less than optimal ethical outcome.

There are at least three models of thought which could be appealed to with regard to the ethical implications of a policy. The first is a consequentialist perspective, in which the benefits a policy might bring are compared to its costs, and a calculation made as to the most cost effective means to achieving the benefits (Duff and Marshall, 2000). An implication of the consequentialist approach is that one must be willing to adopt policies which deprive some individuals of benefits (e.g. privacy or freedom of movement) in order to maximise those of the majority. For example, policies designed to encourage mothers to marry the fathers of their children may have a long term benefit for the child, while producing adverse consequences for the mother (e.g. increased risk of domestic violence, or economic dependence) (Jaffee et al, 2003).

Such concerns would be addressed by the second perspective, which is termed 'side-constrained consequentialism' (Hart, 1968). According to this view, policies which seek efficiently to achieve a consequential good are justified only if they do not conflict with certain non-consequentialist values (Duff, 2004). The precise formulation of these values is open to debate, but principles one might want to apply might include the right to privacy, to self determination, to participation in the decision, to be subject to nothing which harms the individual (even if this is for the good of others). One also needs to take into account any legal constraints or standards, though legal standards on their own are not sufficient to ensure ethical issues are addressed. Although the law codifies much of our moral standards, simply satisfying legal requirements (for example) is not necessarily sufficient to ensure that the outcome is ethical. Thus policies which infringed these agreed rights would not be justified 'even when such violations might efficiently reduce crime' (Duff and Marshall, 2000: 19).

Common to both these perspectives is a distinction between the 'ends' of crime reduction policy and the 'means' used to achieve those ends. Duff and Marshall describe a third perspective, which urges policy makers to treat individuals as responsible moral agents, and sets constraints not only the means to be used, but also on the ends that are to be pursued (Duff and Marshall, 2000). According to this perspective, one must consider more than just the 'end' (e.g. crime reduction), but also the 'how' and 'why'. In other words, a balance has to be struck between the

effective control of crime, and the protection of key principles such as privacy, mobility and freedom of expression and lifestyle (Shaftoe, 2004). Duff and Marshall give the example of a shopping centre's policy to exclude youths so as to reduce crime and increase shoppers' security. While such a policy might meet consequentialist and side-constrained consequentialist criteria, by transforming the public good of crime prevention into a private one by such an exclusionary policy, the end itself has become illegitimate (Duff and Marshall, 2000: 21).

The reality of policy making probably most closely conforms to the side-constrained consequentialist model. A recent Home Office report discussed the various criteria against which potential policy options should be assessed. While maximising the 'public interest' was foremost, it was recognised that a number of constraining factors, including legality, proportionality and the minimisation of unintended consequences had to be taken in to account (Ledbury et al, 2006).

Another way of thinking about the ethics of policy making is to consider the *values* which one wants to underpin the process. In our second report for the Network (*Pathways into Policy*) we discuss the role of the National Institute for Health and Clinical Evidence (NICE) and draw out what we see as the positive aspects of an independent body like NICE for the policy making process. The principles on which NICE bases its decision making also have relevance for the topic of this report (NICE, 2005). Among the factors they take into account in decision making are moral, or ethical, principles. They identify four principles of bioethics (outlined below) which they believe provide a 'simple, accessible and culturally neutral' approach to dealing with the moral issues that arise in healthcare.

The first principle of respect for autonomy places emphasis on individuals' rights to make informed choices about their healthcare. That is not to say that this principle should be universally applied, for as they recognise, factors such as mental or physical incapacity or the fact of finite resources in healthcare, means that a balance may have to be struck between autonomy and other system priorities (NICE 2005:11). It is less clear how the principle of autonomy might be applied in criminal justice policy making, where the subjects of policy (offenders) are those who have transgressed society's norms and can be argued to have surrendered rights such as autonomy as a consequence. However, criminal justice is not just about offenders; victims and those who work within the system are affected by policy as much as offenders, and their autonomy is something which might be considered important when making policy. For example, one could see policies such as victim impact statements, input into sentencing and restorative practices as promoting the autonomy of victims. One can also think of ways in which policy could promote the autonomy of practitioners, for example by encouraging progressive and creative practice at the local level (Smith, 2007).

The second (non-maleficence) and third (benificence) principles are closely related. Non-maleficence asserts that a policy should not seek to inflict damage, or in the words of the medical maxim 'first, do no harm'. In healthcare, of course, most interventions with patients carry the possibility of harm, and again it is the balance between benefit and harm which partly determines the appropriateness of a particular policy (NICE 2005: 11). Again, when one considers policies aimed at offenders, this stricture may seem at odds with criminal justice policies, many of which involve punishment, which may entail an infringement of offenders' liberty. It is generally agreed that the imposition of punishment needs special justification. As Walker states, 'the infliction of something to which a person objects is regarded … as morally wrong unless it can be morally defended' (Walker 1985: 107). That is not to say that punishment necessarily equates to harm. There are many potential aims of

punishment, including rehabilitation and reparation. Thus punishment can result in outcomes for offenders, victims and society at large, which do not inflict harm on the offender (or on others).

The opposite of non-maleficence is beneficence, and requires agents (policy makers) to take positive steps to help others; this goes beyond not harming individuals, or treating individuals autonomously, it calls for the promotion of individuals' legitimate interests (Gericke et al, 2005). Moral philosophers have distinguished between 'positive beneficence' and utility, with the former requiring agents to provide benefits, and the later calling for a balance of benefits, risks and harms (Beauchamp and Childress, 2001). In healthcare, as in criminal justice, it is hard to think of any policy which could be invariably beneficial for everyone, and thus the respect for utility, or balancing benefits and costs, is more appropriate (NICE, 2005:12).

The fourth and final principle which NICE subscribes to is that of distributive justice. This is described as the provision of services in a 'fair and appropriate manner in the light of what is due' (NICE 2005:12). Within this overarching concept there are numerous approaches, including the 'need principle' (resources should be allocated solely on the basis of need), utilitarianism (seeking the greatest good for the greatest number) and egalitarianism (fairness of equality or opportunity) and there is no consensus in healthcare as to which is most appropriate in terms of determining a fair allocation of healthcare resources (NICE 2005:12). One is likely to face similar difficulties in applying this to criminal justice policies, although the basic principle suggests that policies ought to take account of the needs of different groups in society, particularly the most vulnerable ones. For offenders, the trend towards categorising and treating offenders according to the risk they pose could be seen as conforming to this principle.

Whichever model or set of values one adopts, it is clear that there are a number of practical obstacles which must be overcome if the ethical implications of research are adequately to be assessed. In order to assess the utility of a policy, information would be needed on policy alternatives, their likely costs and benefits, both now and in the future. In many cases, the information available may be incomplete or ambiguous. Secondly, it may not always be clear how to judge the benefits and costs of the policy. One is faced, therefore, with a double problem of radical uncertainty about both prediction (as to what the long term consequences might be) and evaluation (as to whether the consequences are on the whole positive) (Bostrum, 2005). This is particularly likely to be the case in relation to pure research such as scopic, as the costs and benefits of various policy options is not part of the remit of the research. In such cases, one has to make 'best guesses' about the likely utility of policy suggestions, and the constraints which may operate.

3.4   Operationalising policy implications

The final aspect of the framework relates to how ideas for policy are operationalised. Taking the example of the impact of parenting on antisocial behaviour, even if one decided that encouraging single mothers to marry the fathers of their children were practical and ethical, one is faced with the question as to how exactly this is to be achieved.

There are a number of 'policy instruments' available to government, including education of the public, provision of information, direct provision of services, economic incentives and regulation or prohibition (Ledbury et al, 2006) and one needs to consider which of these might be effective in addressing a particular

problem. Family policy has often made use of the fiscal system, so one way to encourage marriage may be through the provision of tax breaks, or direct benefits as some US states have done (Trzesniewski et al, 2003). Having identified a policy option, one also needs to consider who should be responsible for funding and / or providing it. In the previous example, taxation is a national domain and as such would need to be funded by and administered through central government, with the policy itself directed at individuals in the target group. If it were shown that the effect of parenting on children's behaviour were also dependent on the community context, then policy makers would need to consider how such resources could be targeted at the right communities (if, indeed, it were possible or legal to do so).

Another consideration in distilling policy implications from research is the fact that policies can be formulated and directed at different structural levels. There are hierarchies in terms of policy scale, policy theme and policy resource target. The terms "policy", "programme" and "project" are often used interchangeably although on closer examination they refer to different points within a hierarchy of planned activity. When considering policy scale, there is a clear distinction between an entire policy domain (e.g. law and order), specific programmes (e.g. the Crime Reduction Programme), initiatives (e.g. the Reducing Burglary Initiative), projects within an initiative (e.g. target hardening), interventions that comprise a project (e.g. alley gates) and tactics that underpin each intervention (e.g. residents' meetings and site appraisals in support of alley gating). These hierarchies are displayed in table 2.

Table 2 Policy Hierarchies
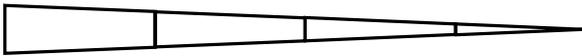
Hierarchy	Broad ▭───────────── Narrow
Policy scale	Policy / Programme / Initiative / Project / Intervention / Tactic
Policy theme	Crime / Violent crime / Robbery / Males / Student robbery
Policy target	Society / Community / School / Peer group / Family / Individual

Policy theme hierarchies are the constituent parts of a distinctive theme within a policy. For example, policies aimed at reducing offending may have components targeting young and older offenders and separate strands working with males and females.

Policy target hierarchies refer to the scale at which resources provided from the policy are allocated. For example, there is the individual, the individual as a member of a household, individuals and households as members of a family or social group (e.g. as defined by ethnicity, religion, class or culture), households, families and groups within a neighbourhood, neighbourhoods/communities within a city and society at large. For youth offending, interventions may be designed and directed towards mechanisms that operate on several scales within the hierarchy. For example, at the *individual level* relevant factors to address through appropriate policy responses would include poor educational attainment, negative attitudes, early anti social behaviour. At the level of the *family*, disruption/poor parenting, abuse and neglect and at *peer level,* attitudes, rejection, emerging 'gang cultures'. Interventions at the level of the *school* might address poor management, class size, exclusions and bullying. Finally, at the *neighbourhood level* relevant influences might be concentrated poverty, social disorganisation and the absence of community cohesion, and the absence of positive role models.

The *interactions* between these different levels may be particularly important in terms of risk and protective factors for offending. If causal mechanisms for offending are identified that clearly demonstrate the combined effect of individual and community level influences, then more innovative policy responses may be required that simultaneously target causal mechanisms operating at the individual and community-level.

The policy implications that stem from empirical research may be relevant for some but not all points in the hierarchy. Thus research findings might inform one or two interventions within a project focused on a few individuals; others may be relevant to an entire programme covering a larger and more heterogeneous population and some results may just be relevant to area-based policies that provide resources shared by individuals as residents of a prioritised neighbourhood.

The levels of resolution used for targeting resources may also be the appropriate scales for detecting possible impacts and outcomes from policy. The search for change potentially attributable to policies may be focused on the individual offender (e.g. who re-offenders or desists over time), the affected community (e.g. for whom the crime rate falls, out migration slows and investment grows) or at other levels (e.g. the institution or service provider whose effectiveness in reducing re-offending grows). Some measures of change only appear at higher levels of resolution where the effects of policies, even if targeted at individuals, can be identified from aggregated data. For example, at the individual level, there are certain attributes of the individual offender (e.g. age and gender) that must be regarded as 'static factors' that are not amenable to change through policy intervention. However, at the community level it is entirely possible for policies to alter the age/gender composition of a neighbourhood. In doing so, they may also alter the predicted relationship between age and gender and crime within that community. Thus age and gender are no longer 'static' from this perspective.

The consideration of policy hierarchy is the final strand to our model, and in the following box we provide a summary of the main features of the model along with some of the questions which need to be addressed by those seeking to influence policy making through research.

<div style="border: 1px solid black;">

Box 1. Key features of the policy implications model

Valid
- Is the research internally valid? Consider issues of:
    - Sampling and recruitment
    - Attrition
    - Defensibility and 'fitness for purpose'
    - Data collection and analysis
    - Reflexivity and objectivity
    - Ethics
- Is the research externally valid? Consider:
    - Sampling method
    - Context of the research
    - Replication
- Construct validity
    - How well are the constructs operationalised?

Feasible
- Are there any identifiable policy implications?
- Are the causal variables manipulable?
- Are they based on data sources not normally available for the population?
- Would the cost/time involved be prohibitive?
- Is there any information on the likely size of benefit?

Ethical
- Would the changes infringe personal freedom/legal rules?
- Would the changes conflict with the values that underpin policy making?
- Any there any likely unintended consequences?

Concrete
- What are the implied policy changes?
- At what level would they operate?
- Who is the target of the policy?
- Who would be responsible for funding/delivery?
- Over what timescale would they be needed?

</div>

We now apply the model to one of the recent publications arising from the scopic research, which looks at the link between reading achievement and anti-social behaviour.

4   CASE STUDY: MATERNAL DEPRESSION AND ANTI-SOCIAL BEHAVIOUR (TRZESNIEWSKI ET AL 2006)

Previous research studies have reported a strong association between reading achievement and antisocial behaviour, but no consensus has yet been reached with regard to the precise causal nature of the relationship. Trzesniewski and colleagues (2006) investigated this relationship by means of a twin study design. We apply the model to their study in the section which follows, and conclude with a discussion of the limitations of our model.

The research (in common with all those conducted under the E-Risk aegis) is a longitudinal study of twins which aims to investigate the relationship between reading achievement and antisocial behaviour. The twin study is a robust methodology which allows the researchers to identify the role of individual and environmental characteristics in the development of antisocial behaviour (Eley and Rijskijk 2005).

There was no threat to internal validity from differential attrition in this study, as each twin was used as the control for his or her twin. Furthermore, the statistical analysis conducted was fully and clearly set out. There were no obvious errors in the analyses conducted, while sample sizes were large.

The authors themselves note that the external validity of the study may be compromised due to the fact it was based on a sample of twins, as mothers of twins may experience different pressures and social contexts to mothers of singletons. However, they demonstrate that the twins in their study had similar reading scores to those of singletons, and argue that the relationship between reading and antisocial behaviour is similar for twins as for singletons. Given that the sample was nationally representative and large, one can be reasonably confident that the results are generalisable to the wider population.

The study set out to investigate the relationship between reading achievement (at age five) and antisocial behaviour (at age seven). The measure used for the former was IQ, on the grounds that as most children have not been taught to read by that age, IQ was the nearest proxy. This could be argued to represent a threat to the construct validity of the study, as IQ measures more than just reading achievement. The second variable, antisocial behaviour, was measured using a series of established instruments of proven validity, so there is no reason to doubt the construct validity of that concept. However, given that the research was seeking, in part, to uncover the reasons for involvement in crime, the question arises of whether antisocial behaviour at age seven (however measured) is a valid predictor of behaviour which would be criminal once the age of criminal responsibility (10) is passed. This is not directly addressed by the authors, and means that questions remain as to the strength of the construct validity of the study.

The implications of the study are that by improving reading ability early in a child's life, later antisocial behaviour can be mediated. The variable of reading ability is something which it is clearly possible to change. It would also appear feasible to do so, given that reading ability is one of the key skills which schools both measure and seek to improve. Given the association between high reading ability and positive outcomes such as employment and reduced involvement in crime, there are no obvious ethical reasons not to seek to improve reading ability. There might be ethical concerns, however, if it was decided that the policy should be implemented only with certain groups (e.g. those in most deprived wards) as this may conflict with the principle of egalitarianism. On the other hand, it could be argued that

consequentialist principles would recognise that strict equality may not necessarily lead to the optimal outcome for society. The final step in applying the model is to consider by what means the policy implications might be achieved.

Reading ability is unlikely to be improved through policy tools such as provision of information, regulation or prohibition. As a public good, the improvement of reading ability is best achieved through the direct provision of education, for example by incorporating into the curriculum a stronger emphasis on reading, and making use of research into what are the most effective strategies to achieve this. Not all education takes place in schools, of course, and reading ability might also be improved by encouraging mothers to read to and interact more with their babies and toddlers, perhaps providing them with free or discounted books or through other economic incentives.[8] Decisions would need to be made regarding the optimal level of reading ability. For example, should the education system seek to improve the reading ability of the entire age group by a certain percentage, or to ensure that everyone attains at least a certain minimum level? In answering this question, policy makers might require additional information on the type of relationship between reading ability and crime. For example if it was a simple linear one then the former approach might be adopted. If however, there was a 'tipping point', such that once a particular level of reading ability was attained the involvement in anti-social behaviour declined significantly, then the latter would be more appropriate. It was not clear from the study the precise nature of the relationship, but the authors did find that the relationship was stronger for boys than for girls, suggesting that resources would best be focused on them.

Their findings also suggest that it would be more effective to intervene to improve reading ability in the early years of schooling, before patterns of behaviour become entrenched. Reading ability is unlikely to be randomly distributed across schools – in some schools the majority of pupils will have lower than average achievement, while in others most pupils will score highly. Rather than applying the policy to every school, therefore, it might be more cost effective to focus on schools with a disproportionate number of pupils with low reading ability.

We have sketched out how a model for deriving policy implications can be applied to empirical research. The example we chose had strong internal and external validity, although there was some question mark over the construct validity of the measures used. Reading ability, which was shown to be associated with antisocial behaviour, is a variable which it is possible and feasible to change, and there were no obvious ethical obstacles to such a policy. Some policy options were suggested to achieve this. In the next section we consider some limitations of the model.

---

[8] Policy makers in New York have gone down this route, offering parents financial incentives for carrying out basic parenting activities such as discussing their child's school tests. See: http://www.nyc.gov/html/ceo/html/programs/opportunity_nyc.shtml

5 LIMITATIONS

The first limitation relates not so much to the model itself as to its application to real world research. Users of research, be they policy makers or academics, do not have access to 'the study' itself. The study does not exist for others until it is reported. A major difficulty in trying to derive lessons for policy from published research is the possibility that key methodological aspects of the study will be inadequately reported.[9] In the example of research discussed above, the published account of the research was well reported, so it was possible to assess validity, practicality and feasibility of the policy implications. Not all studies are likely to be reported so fully and the question arises as to what to do in such cases. One solution would be to contact the authors directly in order to elicit any missing information, thus ensuring that the assessment is based on the most accurate representation of the research as conducted. However, this might not always be practical, especially if policy makers or civil servants wished to make decisions quickly on the basis of a range of published evidence. In which case, the problem for those reading research is whether to treat the absence of reporting of an aspect of methodology (e.g. attrition) as evidence that it was not a problem. This dilemma has faced designers of rating scales which attempt to assess methodological rigour. In one scale for medical trials (the PEDro scale), for example, credit is only given when a criterion is clearly satisfied. Thus, if on a literal reading of the report it was possible that a criterion was not satisfied, a point would not be awarded for that criterion; what they term the 'guilty until proven innocent' approach (Verhagen et al, 1998). In other scales, however, the benefit of the doubt is sometimes given. Downes and Black (1998) advise raters, for example, that if the distribution of the data is not described, to assume that the statistical tests used were appropriate.

Whichever approach is taken, there is a danger that one is rating the quality of a study as can best be determined from the published material, rather than the quality of the research as conducted, and therefore one risks making an incorrect assessment of its rigour. Thus, the internal validity of a study may be under-estimated if the authors fail to mention key features such as rates of attrition or provide inadequate details of statistical analysis. We have decided to adopt the prudent 'innocent until proven guilty' approach on the basis that it is more reasonable to favour the assumption that if a study was carefully conducted then it will also be carefully reported, over the assumption that a carefully conducted study may be poorly reported (Wilcox, 2005).

Secondly, the model will be of more utility to some types of research than others. Research that does not have a precise policy focus is unlikely automatically to lead to policy change. This might be the case for a number of reasons. The researchers may not be sufficiently familiar with the policy making process – there may be clear lessons for policy but the conclusions to the study fail to spell out the policy implications of the research. Alternatively, it might simply be beyond the remit of the study to do so. Thus if the primary focus of the research is on the measurement and/or explanation of criminal offending, victimisation or offence attributes the findings will not automatically indicate what steps should be taken to impact upon crime.

Turning to the model itself, although it presents a defensible, structured framework for deriving policy implications from research, the process inevitably calls for decision making, involving subjective judgements, for example, as to whether plausible threats

---

[9] This is why some have argued for the inclusion of descriptive validity (see above) as part of the overall assessment of the methodological quality of a study (Lösel and Koferl, 1989).

to validity have been ruled out, or whether the variables identified could feasibly be manipulated by means of public policy. In exercising such judgement one is bound to draw upon one's knowledge of the subject area, rather than relying solely on the conclusions or implications drawn by the researchers themselves. In so doing, one introduces the possibility of unreliability, in that two people with varying levels of background knowledge reading the same study may arrive at different conclusions as to its policy implications. Ideally, in order to promote greater reliability it would be advisable for assessments of research to be conducted separately by two or more people, and for any disagreements to be discussed. In order to help resolve such differences of opinion, it would useful to record the reasoning behind the decisions taken at each stage of the model. It is recognised that in practice it may be unrealistic for more than one person to assess the research, and one may have to accept some unreliability in the interests of ease of use and efficiency.

Finally, the model represents a significant simplification of reality. If one considers the policy making process, there are clearly more factors involved than just the results of empirical research, as we discuss in our second report. The idealised 'rational model' of policy making involves a number of steps, including the verification and definition of the policy problem, establishment of evaluation criteria, identification and evaluation of alternative policies, selection among alternative policies and monitoring of policy outcomes (Haas and Springer, 1998). Despite the rhetoric of evidence-based policy making (e.g. Blunkett 2000), the 'evidence' is just one influence on the criminal justice policy process. That is not to say that evidence *should* have a primary role in all areas of public policy. Indeed Young and Sanders argue that policy 'should ultimately be led by values, not evidence' and that the role of evidence is to test the 'factual assumptions on which value choices are (usually) based' (Young and Sanders, 2003: 340). Other factors which policy makers might legitimately take into account are cost, public opinion and legal constraints. More important is the fact that many of the problems facing government are not of the kind amenable to the simple application of the 'rational model'. As Howlett and Ramesh have argued, the reality is a more complex phenomenon which falls somewhere between the purely rational and the purely political (Howlett and Ramesh 2003).

There are many reasons for this, including the promotion of vague, ambiguous or even contradictory policies, the lack of viable alternative policies, the commissioning of evaluation after a policy has been instituted, and the difficulties of measuring the success of policies designed to deal with socially constructed problems (Haas and Springer 1998). This has implications for how the model is used and it may be that the process of deriving policy implications does not necessarily follow in the order described above. For example, it may be necessary to operationalise the implications before it can be decided whether or not they are feasible or desirable.

As recognised above, policy making is rightly a political process, and it is perhaps more realistic to expect policy to be 'evidence-aware' or 'evidence-influenced' (Nutley et al 2002: 2). However, as many writers have argued, even this more modest expectation is not being met, as policy makers ignore or selectively use the evidence (Burnett and Appleton, 2004; Wilcox, 2003). In criminal justice, this can have the disastrous consequence that policies are introduced, such as 'boot camps' (Nathan 1995), in the face of clear evidence that they do not work.

6   CONCLUSIONS

The current era of 'evidence-based' policy making has been characterised by a significant increase in funding for criminological research. Given this, it has been increasingly important for policy makers and practitioners to be able to make use of the findings of criminological research into the causes of crime. The framework we have outlined here is designed to help with the aim of deriving policy implications from research. Further testing and use of the framework will be necessary before deciding whether it proves a useful and practical tool. Of course, it is not enough simply to apply mechanistically a model such as we have outlined. As we discuss in our second report, there are many obstacles placed before those who wish research to have more influence on policy. One of the key ones is that of language. Policy recommendations (whether proposed by those conducting research or by users of our model) need to be written in a language that policy practitioners, politicians and local communities can identify with and understand. They need to be unambiguous and the steps required to implement them need to be feasible given budgetary constraints and the capacity of agencies to deliver them. They also need to be ethically sound and realistic in terms of the speed and scale of their implementation, the reaction and possible opposition from affected communities and the rigours of political scrutiny.

The framing of policy recommendations is a highly skilled job informed by an understanding of the decision making process and a shrewd assessment of the chances of acceptance by those in positions of power and control over resources. These are skills that emanate from experience-based knowledge rather than academic research training. Although the latter may equip researchers in quantitative and qualitative research techniques it will not necessarily enable them to identify the action implications of research or to express these in the language needed to communicate these to diverse audiences. Thus it is unrealistic to assume that conducting empirical research equips the investigator with the necessary skills for distilling 'actionable' policy recommendations from research findings.

The remedy to this dilemma is to incorporate, within the empirical research of the 'problem' (e.g. juvenile offending), an analysis of the decision making processes in relevant policy bodies (i.e. those agencies that 'respond' to offending). If the operation, culture, resources, priorities and decision making arrangements of organisations that respond to offending are included in the research alongside that on definitions, measurement and likely causal mechanisms, the chances of identifying clear policy options will be that much greater.

The *types* of policy implication that arise from empirical research largely depend on the focus of the research. If the research is essentially of the problem diagnosis variety (analysis for policy), the findings may provide the conceptual model or theory of change that identifies the general direction or thrust of the policy. If the research examines alternative modes of service provision, the implications are more likely to inform the management and policy implementation process. If the research examines the inter-relationships between policy strands currently operational within a neighbourhood (policy analysis) the implications may inform how the delivery of policy and targeting of investment might be improved through better policy coordination, rationalisation and improved social or spatial targeting.

Empirical research that aims to uncover possible causal processes that place individuals at greater risk of victimisation or of being an offender will not necessarily examine the implication of the findings for policy design, implementation and coordination. The chances are that these questions would not be addressed.

To sum up, we have seen that there are two main obstacles to the application of the framework, the first being the inadequate reporting of key aspects of methodology in some studies, and the second being the deviation from an evidence-based, or even evidence-aware approach to policy making. It is hoped that by offering a model to policy makers and others to use, that it will encourage researchers to be more rigorous when writing up issues of validity and policy design and that policy makers will be inclined to place more weight upon the findings of criminological research.

REFERENCES

Baldwin, J. and Bottoms, T. (1976) *The Urban Criminal* London: Tavistock.

Beauchamp, T.L. and Childress, J.F. (2001) *Principles of Biomedical Ethics* Oxford: Oxford University Press.

Binder, A. and Meeker, J. (1992) 'Implications of the Failure to Replicate the Minneapolis Experiment Findings' American Sociological Review 58(6): 886-888.

Blunkett, D. (2000) *Influence or Irrelevance Can Social Science Improve Government?* Secretary of State's ESRC Lecture 2 February 2000 London: DfEE.

Bostrum, N. (2005) *Status Quo Bias in Bioethics: The Case for Human Enhancement* Paper Presented at Oxford-Scandinavia Ethics Summit 13-15 July, Oxford, UK.

Brounstein, P. J. Emshoff, J. G. Hill, G. A. and Stoil, M. J. (1997) 'Assessment of Methodological Practices in the Evaluation of Alcohol and Other Drug (AOD) Abuse Prevention', *Journal of Health and Social Policy* 9(2): 1-19.

Burnett, R. and Appleton, C. (2004) 'Joined-up Services to Tackle Youth Crime: A Case-Study in England' *British Journal of Criminology* 44(1): 34-54.

Cann, J. Falshaw, L. Nugent, F. and Friendship, C. (2003) *Understanding What Works: Accredited Cognitive Skills Programmes for Adult Men and Young Offenders* Home Office Research Findings 226 London: Home Office.

Chalmers, T. Smith, H. Blackburn, B. Silverman, B. Schroeder, B. Reitman, D. and Ambroz, A. (1981) 'A Method for Assessing the Quality of a Randomized Control Trial', *Controlled Clinical Trials* 2: 31-49.

Codd, J.A. (1988) 'The Construction and Deconstruction of Educational Policy Documents' *Journal of Education Policy* 3(3): 235-247.

Colebatch, H. (2005). Policy Analysis, Policy Practice and Political Science. *Australian Journal of Public Administration, 64*(3), 14-23.

Coleman, J. (1975). Problems of Conceptualization and Measurement in Studying Policy Impacts. In K. Dolbeare (Ed.), *Public Policy Evaluation*. London: Sage.

Cook, T. D. and Campbell D. T. (1979) *Quasi-experimentation: Design and Analysis Issues for Field Settings* Chicago: Rand McNally College.

Cullen, F. T. and Gendreau, P. (2000) 'Assessing Correctional Rehabilitation: Policy, Practice, and Prospects' in J. Horney (ed) *Policies, Processes and Decisions of the Criminal Justice System: Volume 3* Washington, DC: National Institute of Justice.

Davies, P. (2004) *Is Evidence Based Government Possible?* Jerry Lee Lecture 4th Annual Campbell Collaboration Colloquium Washington DC.

Downs, S. H., and Black, N. (1998) 'The Feasibility of Creating a Checklist for the Assessment of the Methodological Quality Both of Randomised and Non-randomised Studies of Health Care Interventions', *Journal of Epidemiology and Community Health* 52: 377-384.

Duff, A. (2004) 'Legal Punishment' in Zalta, N. (ed.) *The Stanford Encyclopedia of Philosophy (Spring 2004 Edition)* Available online at: http://plato.stanford.edu/archives/spr2004/entries/legal-punishment/

Duff, R. and Marshall, S. (2000) 'Benefits, Burdens and Responsibilities' in Von Hirsch, A., Garland, D. and Wakefield, A. *Ethical and Social Perspectives on Situational Crime Prevention* Oxford: Hart Publishing.

Edwards, T. (2000) '"All the Evidence Shows…": Reasonable Expectations of Educational Research', *Oxford Review of Education* 26(3/4): 299-311.

Ekblom, P. (2002) 'From the Source to the Mainstream is Uphill: The Challenge of Transferring Knowledge of Crime Prevention through Replication, Innovation and Anticipation' in N. Tilley (ed) *Analysis for Crime Prevention* Cullompton: Willan.

Eley, T. and Rijsdijk, F. (2005) 'Introductory Guide to the Statistics of Molecular Genetics' *Journal of Child Psychology and Psychiatry* 46(10): 1042-1044.

Farrington, D. P. (2002) *Methodological Quality Standards for Evaluation Research* Paper Presented at the Third Annual Jerry Lee Crime Prevention Symposium, University of Maryland.

Farrington, D. P. (1992) 'Explaining the Beginning, Progress and Ending of Antisocial Behaviour from Birth to Adulthood' in J. McCord (ed.) *Facts, Frameworks and Forecasts: Advances in Criminological Theory*, iii, New Brunswick: Transaction.

Friendship, C. Blud, L. Erikson, M, Travers, R. and Thornton, D. (2003) 'Cognitive-behavioural Treatment for Imprisoned Offenders: An Evaluation of HM Prison Service's Cognitive Skills Programmes', *Legal and Criminological Psychology* 8: 103-114.

Gericke, C.A., Riesberg, A. and Busse, R. (2005) 'Ethical Issues in Funding Orphan Drug Research and Development' *Journal of Medical Ethics* 31: 164-168.

Gibbs, L. (1989) 'Quality of Study Rating Form: An Instrument for Synthesizing Evaluation Studies', *Journal of Social Work Education* 25*: 55-66.

Gill, M., Rose, A,. Collins, K. and Hemming, M. (2006) 'Redeployable CCTV and Drug-related Crime: A Case of Implementation Failure' *Drugs: Education, Prevention and Policy* 13(5): 451-460.

Gordon, I., Lewis, J. and Young, K. (1977) 'Perspectives on Policy Analysis' *Public Administration Bulletin* 25: 26-35.

Granger, R. 1998 Establishing Causality in Evaluations of Comprehensive Community Initiatives in K. Fulbright-Anderson, A. Kubish, and J. Connell (eds.) *New Approaches to Evaluating Community Initiatives: Theory, Measurement and Analysis* vol. 2, Queenstown, The Aspen Institute

Guba, E.G. and Lincoln, Y.S. (1989). *Fourth generation evaluation*. California: Sage Publications.

Haas, P., and Springer, J. (1998). *Applied Policy Research: Concepts and Cases*. London: Garland Publishing.

Hart, H.L.A. (1968) *Punishment and Responsibility* Oxford: Oxford University Press.

Howlett, M., and Ramesh, M. (2003). *Studying Public Policy: Policy Cycles and Policy Subsystems*. Oxford: Oxford University Press.

Jaffee, S. R., Moffitt, T. E., Caspi, A., and Taylor, A. (2003) 'Life With (or Without) Father: The Benefits of Living with Two Biological Parents Depend on the Father's Antisocial Behavior' *Child Development* 74: 109-126.

Ledbury, M., Miller, N., Lee, A., Fairman, T. and Clifton, C. (2006) Understanding Policy Options Home Office Online Report 06/06 Available online at: http://www.homeoffice.gov.uk/rds/pdfs06/rdsolr0606.pdf

Logan, C. (1972) 'Evaluation Research in Crime and Delinquency: A Reappraisal', *Journal of Criminal Law, Criminology and Police Science* 63(4): 378-387.

Lösel, F. and Koferl, P. (1989) 'Evaluation Research on Correctional Treatment in West Germany: A Meta-Analysis' in F. Wegener, F. Lösel and J. Haisch (eds) *Criminal Behavior and the Justice System: Psychological Perspectives* New York: Springer-Verlag.

May, C. (1999) *Reconviction Following a Community Sentence: Explaining the Role of Social Factors* Home Office Research Study 192 London: Home Office.

Moffitt, TE, Caspi, A, Rutter, M and Silva, PA (2001) *Sex Differences in Antisocial Behavior: Conduct Disorder, Delinquency, and Violence in the Dunedin Longitudinal Study* Cambridge, UK: Cambridge University Press.

Moher, D. Jadad, A. R. Nichol, G. Penman, M. Tugwell, P. and Walsh, S. (1995) 'Assessing the Quality of Randomized Controlled Trials', *Controlled Clinical Trials* 16: 63-73.

Nathan, S. (1995) *Boot Camps: Return of the Short, Sharp Shock*. London: Prison Reform Trust.

NHS Centre for Reviews and Dissemination (2001) *Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidelines for Those Carrying out or Commissioning Reviews* (2nd ed.) York: York Publishing Services.

NICE (2005) *Social Value Judgements: Principles for the Development of NICE Guidance* London: NICE. Available online at: http://www.nice.org.uk/download.aspx?o=283494 (accessed 12 September 2007)

Nutley, S, Davies, H. T. O. and Walter, I (2002) *Evidence-Based Policy and Practice: Cross Sector Lessons from the UK* Paper Presented at the Social Policy Research and Evaluation Conference Wellington, New Zealand.

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

Pawson, R. and Tilley, N. (1994) 'What Works in Evaluation Research' *British Journal of Criminology* 34(3):291-306.

Plewis, I. (2000) 'Educational Inequalities and Education Action Zones' in C. Pantazis and D. Gordon (eds) *Tackling Inequalities: Where We Are Now and What Can be Done* Bristol: Policy Press.

Rosenbaum, D. (2002) 'Evaluating Multi-agency Anti-crime Partnerships: Theory, Design and Measurement Issues' in N. Tilley (ed.) *Evaluation for Crime Prevention* Crime Prevention Studies Volume 14 Cullompton: Willan Publishing.

Sanderson, I. (2000) 'Evaluation in Complex Policy Systems' *Evaluation* 6(4): 433-454.

Shaftoe, H. (2004) *Crime Prevention: Facts, Fallacies and the Future* Basingstoke: Palgrave.

Sherman, L.W. and Berk, R. (1984) *The Minneapolis Domestic Violence Experiment* Police Foundation Report No. 1 Washington DC: Police Foundation.

Sherman, L. W. and Gottfredson, D. C. (1998) 'Appendix: Research Methods' in L. W. Sherman, D. Gottfredson, D. MacKenzie, J. Eck, P. Reuter and S. Bushway (eds) *Preventing Crime : What Works, What Doesn't, What's Promising* Washington, DC: U.S. Dept. of Justice, Office of Justice Programs. Available online at http://www.ncjrs.org/works/ (checked 10 May 2005).

Sherman, L. W. Gottfredson, D. MacKenzie, D. Eck, J. Reuter, P. and Bushway, S. (1998) *Preventing Crime: What Works, What Doesn't, What's Promising* Washington, DC: U.S. Dept. of Justice, Office of Justice Programs. Available online at http://www.ncjrs.org/works/ (accessed 30 August 2007).

Smith, R. (2007) *Youth Justice: Ideas, Policy, Practice* (2nd ed). Cullompton: Willan Publishing.

Spencer, L., Richie, R., Lewis, J. and Dillon, L. (2003) Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence London: Cabinet Office. Available online at: http://www.policyhub.gov.uk/docs/qqe_rep.pdf (accessed 3 September 2007).

Stevenson, C. and Cooper, N. (1997) 'Quantitative and Qualitative Research' *The Psychologist* 10(4):159-160.

Tenbensel T (2004) 'Does more evidence lead to better policy? The Implications of Explicit Priority-Setting in New Zealand's Health Policy for Evidence-based Policy' *Policy Studies* 25(3): 189-207.

Tilley, N. and Laycock, G. (2000) Joining Up Research, Policy and Practice about Crime *Policy Studies* 21(3): 213-227.

Trzesniewski, K., Moffitt, T. and Caspi, A. (2006) 'Revisiting the Association Between Reading Achievement and Antisocial Behavior: New Evidence of an Environmental Explanation from a Twin Study' *Child Development* 77(1): 72-88.

Trochim, W. M. (2000) The Research Methods Knowledge Base (2nd ed.) Available online at: http://www.socialresearchmethods.net/kb/

Verhagen, A. P. de Vet, H. C. de Bie, R. A. Kessels, A. G. Boers, M. Bouter, L. M. and Knipschild, P. G. (1998) 'The Delphi List: a Criteria List for Quality Assessment

of Randomized Clinical Trials for Conducting Systematic Reviews Developed by Delphi Consensus', *Journal of Clinical Epidemiology* 51(12): 1235-1241.

Walker, N. (1985) *Sentencing Theory, Law and Practice* London: Butterworths.

Wikström, P.O.H. (1991) *Urban Crime, Criminals, and Victims: the Swedish Experience in an Anglo-American Comparative Perspective*. Stockholm: Springer-Verlag.

Wikström, P.O. (2005) Cambridge Network for the Study of the Social Contexts of Pathways in Crime: Annual Report for Period January 2004 to December 2004. Cambridge: Institute of Criminology.

Wilcox, A. and Hirschfield, A. (2007) *Pathways into Policy: A Study of the Relationship between Research, Policy and Government* Huddersfield: Applied Criminology Centre.

Wilcox, A. (2005) *The Validity of Reconviction Studies* (Unpublished PhD Thesis) Oxford: University of Oxford.

Wilcox, A. (2003) 'Evidence-based Youth Justice? Some Valuable Lessons from an Evaluation for the Youth Justice Board', *Youth Justice* 3(1):19-33.

Young, R. and Sanders, A. (2003) 'The Foresters' Dilemma: The Influence of Police Research on Police Practice' in L. Zedner and A. Ashworth (eds) *The Criminological Foundations of Penal Policy: Essays in Honour of Roger Hood* Oxford: Oxford University Press.