

Criminal justice evaluation in the UK



EUROPEAN
UNION



European Social Fund
Investing in people



OPAC. Experts in action



CENTER FOR
THE STUDY OF
DEMOCRACY



Manchester
Metropolitan
University

The project is implemented with the financial support of
Operational Programme "Administrative Capacity", co-financed by the European Union through the European Social Fund

This material is owned by Chris Fox and Dan Ellingworth

Speakers

Chris Fox

- Principal Lecturer in Criminology at Manchester Metropolitan University
- Home Office Regional Advisor on Research and Analysis for the North of England
- Joint Editor of Safer Communities – a practitioner journal with a focus on crime reduction
- Directed numerous evaluations of crime reduction and criminal justice projects and programmes for the UK Home Office, UK Ministry of Justice and UK Department of Health.
- Worked extensively with local crime reduction agencies including police, probation, drug services and youth justice.

Dan Ellingworth

- Senior Lecturer in Criminology at Manchester Metropolitan University
- Teaching interests in criminological theory and research, youth justice, victimisation and research methodology.
- Research experience on the analysis of the British Crime Survey, UK Census datasets, and a variety of other large government datasets.
- Worked on a range of research projects addressing repeat victimisation, geographical distribution of crime, domestic violence, education and crime, policing and criminal justice interventions.

Outline

1. Why evaluate?
 - Accountability
 - Evidence-based policy
2. What to evaluate?
 - Factors limiting the potential for successful evaluation
3. When to evaluate?
 - Ex ante evaluation
 - Ex poste evaluation
 - Systematic reviews
4. Who should do the evaluation?
 - Commissioning
 - Internal evaluators
 - External evaluators
 - Project management
5. How to evaluate
 - Should it work? (Theories of change)
 - Can it work? (Implementation evaluation)
 - Does it work? (Impact evaluation)
 - Is it worth it? (Economic evaluation)
 - Data for evaluation
6. Where to find out more

Why evaluate?

Overview

- Evaluation can contribute to a process of ensuring accountability, but is not a substitute for appropriate monitoring, audit and review.
- Evaluation is a crucial element of a modern, evidence-led approach to policy making.

Accountability

- Accountability will be ensured through:
 1. ongoing performance management / monitoring
 2. audit (financial and non-financial)
 3. review / inspection, and finally
 4. evaluation
- Evaluation will draw on data generated by the 1st three – particularly monitoring data
- Evaluation is not a substitute for the 1st three

'Modern policy making'

- **Forward looking: Defining policy outcomes and taking a long term view**
- Outward looking: Take account of national, European and international situation; learning from experience of other countries; recognising regional variations.
- Innovative, flexible: Questioning established ways of dealing with things, encouraging new and creative ideas, identifying and managing risk.
- Joining up: Joining up the work of different government departments; ensuring that implementation is part of the policy process.
- Inclusive: Consulting those responsible for implementation and those affected by the policy; carrying out an impact assessment
- **Evidence based: Basing policy decisions and advice upon the best available evidence; ensuring evidence is available in an accessible and meaningful form.**
- **Evaluated: Systematic evaluation of the effectiveness of policy.**
- **Reviews: Policy constantly reviewed to ensure it is dealing with problems it was designed to solve.**
- **Lessons learned: Learning from experience of what works and what does not**

Cabinet Office (1999) *Professional policy making for the twenty-first century* London: Cabinet Office

What is evidence-based policy?

- Evidence-based policy (EBP). An approach that:
 - “helps people make well informed decisions about policies, programmes and projects by putting the best available evidence from research at the heart of policy development and implementation” (Davies 1999 quoted in Davies 2004)
- Contrasted with opinion-based policy which:
 - “relies heavily on either the selective use of evidence (e.g. on single studies irrespective of quality) or on the untested views of individuals or groups, often inspired by ideological standpoints, prejudices, or speculative conjecture.” (Davies 2004)

What to evaluate?

Overview

- The UK (and US) experience provides evidence of a wide range of interventions and initiatives being evaluated.
- Key limitations to evaluation include accessing research subjects, reliability of and access to data, intangible outcomes and opportunities to implement robust effectiveness studies.

Range of evaluation

- Policing
 - Tendency to focus on new initiatives/projects rather than the efficacy of mainstream policing.
- Local crime reduction initiatives
 - Strong tradition of evaluating situational initiatives , but less success in successfully evaluating social/community initiatives
- Courts
 - Research and evaluation on mainstream court processes are limited. Evaluations of specialist courts underway, but currently a much stronger evidence-base in the US
- Drug treatment
 - Relatively well evaluated area where robust studies of effect have been undertaken, influenced in part by the health sector where there is a stronger tradition of robust impact evaluation in place

Range of evaluation

- Prisons
 - Robust evaluations of mainstream, custodial disposals are relatively rare although recent policies around reducing re-offending and commissioning are encouraging more evidence-based sentencing and structures are being put in place to facilitate this.
- Community sentences
 - A number of evaluations of different community sentences and programmes for addressing specific types of offending behaviour have been undertaken, although robust evaluations of effect are limited.

Range of evaluation

- Drug enforcement
 - A growing body of evidence on the operation of illicit drugs markets. Few evaluations, but some econometric work on drug market levers.
- Organised crime
 - Some work relating to localised 'gang-type' interventions in the US and UK
 - The trans-national and covert nature of much organised crime makes quantitative assessments difficult

Accessing research subjects

- Accessing stakeholders is a challenge in any evaluation.
- In the criminal justice sector particular issues occur in relation to:
 - Offenders: Most likely to be successfully engaged when evaluators work with the Probation Service
 - Serving prisoners: Very difficult to gain access to serving prisoners (see Case Study 3).
 - Victims: Victims of violence and sexual offences require respectful, skilled and sensitive interviews.
 - The judiciary: Access strictly controlled.

Reliability of and access to data

- Key data sets provide rich source of data on incidence of crime, victim experience, criminogenic factors and re-offending.
- Data is generated by Criminal Justice Service (CJS) agencies and through surveys.
 - CJS data can provide detailed information at the level of the individual or incident but only captures crime known to the CJS, is subject to changes in reporting practice, varying quality and issues of access.
 - Survey data is better able to capture unreported crime, but is not able to identify local trends.

Robust effectiveness studies

- Common challenges in undertaking robust effectiveness studies include:
 - Defining impact measures: Choice of impact measures is often limited by availability of data or resources to collect primary data.
 - Identifying a comparator: A comparator group or area may not be readily available.
 - Throughputs: Projects often struggle to generate sufficient throughputs to allow for robust statistical analysis.
 - Timescales: Often evaluation timescales preclude appropriate follow-up periods to identify the long-term impact of interventions.

When to evaluate?

Overview

- Evaluation can support all elements of the policy-making cycle.
- Ex ante evaluation is seen as increasingly important in the development of evidence-based policy.
- To maximise its potential, ex poste evaluation should be designed into the policy-making cycle from an early stage.

When to evaluate?

- The intervention is innovative (either because it is doing something that has not been tried before, or because it is doing something that has been tried before but is doing it in a new context).
- The intervention is high-risk (here evaluation might be used to provide an early indication of the likely effects of the intervention).
- The intervention is costly (almost by definition this probably makes the project relatively high-risk and therefore a possible candidate for evaluation)

Questions to ask

Issue	Questions to ask prior to starting the evaluation
The intervention	Are the interventions, and the target population, clear and identifiable?
	Are the outcomes clear, specific and measurable?
	If the intention of the evaluation is to address issues of impact or efficiency, is the intervention sufficiently mature for this to be practical?
Evaluation users	Is it clear who the ultimate users of the evaluation will be?
	What are the needs of the ultimate users (for example, do they have to make a specific decision at a specific point in time and how robust does the evidence for this decision have to be)?
Evaluation stakeholders	Do significant evaluation stakeholders agree on the mission, goals of other critical issues for the intervention being evaluated? If not will it be possible to design an evaluation that addresses their various needs, concerns, political values or ideologies?
	If the intervention is being implemented in a multi-agency context are all the relevant agencies likely to cooperate with the evaluators?

Questions to ask

Issue	Questions to ask prior to starting an evaluation
Evaluation questions	Is it clear what questions the evaluation is required to answer and can a well-designed and properly resourced evaluation answer them?
	Is it possible that there are hidden agendas?
Methods and procedures	Can useful, timely and credible information be gathered for this evaluation?
Change over time	Will the nature of intervention change during the course of the evaluation and if so what impact might this have on the design of the evaluation?
	Will personnel involved in the intervention change and if so will new personnel still cooperate with the evaluation and still have sufficient insight into the intervention?
	Will the evaluation users change during the course of the evaluation and if so will they still require the same outputs from the evaluation?

Formative and summative

Formative

- A formative evaluation asks how, why, and under what conditions does an intervention work, or fail to work?

Summative

- A summative evaluation asks questions about the impact or effects of an intervention. This will normally extend to the total impact of the intervention, not just the extent to which stated goals were achieved.

“The distinction between summative and formative evaluations is not always as rigid as the above characterisation might suggest. Determining whether or not a policy has had an impact often involves asking questions about **how** it has done so, **for whom, why,** and **under what conditions** it has had/not had the effect.”

Ex poste and ex ante

ex post evaluation

- “an evaluation conducted either on or after completion of an intervention” .
- Most effective when designed into a project or programme at an early stage.

ex ante evaluation

- “an evaluation conducted before the implementation of an intervention”
- Most commonly used at a policy-making level.
- European Commission places emphasis on the importance of ex ante evaluation:
 - "If you don't know where you're going, how will you ever know if you get there?"

Who should do the evaluation?

Overview

- In the UK most government departments and executive agencies have a research department.
- Many government departments maintain a distinction between 'policy customers' and 'research providers'.
- A substantial proportion of research is contracted out.

Government research in UK

- Most government departments and executive agencies have a research department.
- Most departments operate a distinction between 'policy customers' and 'research providers'.
- The Government Social Research service provides cross-departmental support for research and evaluation
- Research staff in departments may be affiliated with cross-departmental specialist groupings eg economics, operational research.

Commissioning evaluation

- Most large evaluations on crime reduction and criminal justice in the UK are commissioned by the Research, Development and Statistics Department of the Home Office, the Research Department at the Ministry of Justice and the Research Department at the Youth Justice Board.
- Competitive tendering is the norm and framework agreements for the provision of particular services are common.
- Local Crime and Disorder Reduction Partnerships and Youth Offending Teams will commission small evaluation projects.

Evaluation providers

- A mixed economy exists:
 - Private sector eg Ecotec, Matrix Knowledge Group, BMRB
 - Not for Profit eg Universities, Nacro, Crime Concern
 - Hybrids eg National Centre for Social Research is 'not for profit' but operates along commercial lines.

How to evaluate

Overview

- Clarity is required in what sense “working” is meant – evaluations can provide a range of answers to different questions.
- Different evaluation methodologies are better suited to different evaluation questions.

An evaluation framework

- **Should it work? (Theory of change)**
 - What is the underlying ‘theory of change’ which explains why the intervention will make an impact?
- **Can it work? (Implementation evaluation)**
 - Has the project been properly implemented? What were the challenges to implementation and how were they overcome?
- **Does it work? (Impact evaluation)**
 - What is the impact of the intervention?
- **Is it worth it? (Economic evaluation)**
 - What are the resource implications of implementing the intervention and what benefits will it deliver?

Overview of possible methods

Economic analysis				•
Surveys of those affected			•	
Interviews with service users	•	•	•	
Capture / analysis of case files or other administrative data		•	•	
Capture / analysis of monitoring data		•		
Observation of key processes		•		
Staff interviews		•		
Interviews with key stakeholders	•	•		
Review of international research literature	•			
Review of project literature	•			
Should it work?	•			
Can it work?				
Does it work?				
Is it worth it?				

Should it work? Theories of change

- Sometimes programmes and projects:
 - don't have clear aims or change their aims over time
 - are 'shoe horned' into a set of funding criteria to access that funding
 - don't have clearly defined processes and structures
 - are implemented in a multi-agency context where different agencies will have different ideas about what the project is trying to achieve

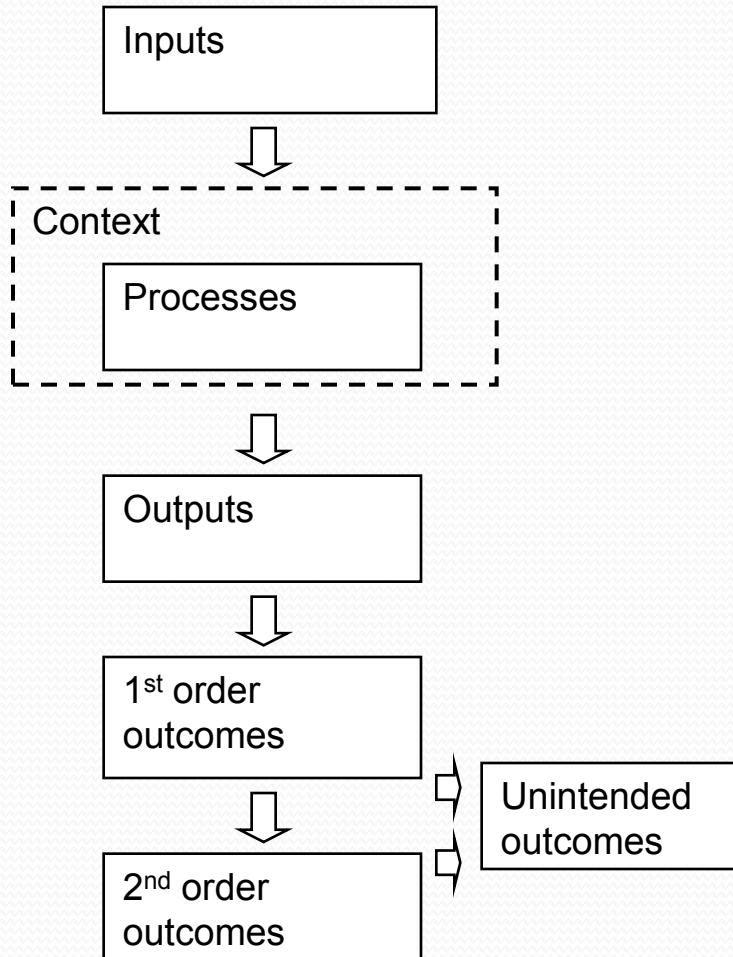
Theories of change

- “What is the conceptual link from an intervention's inputs to the production of its outputs and, subsequently, to its impacts on society in terms of results and outcomes?”

(United Kingdom Evaluation Society Glossary of evaluation terms http://www.evaluation.org.uk/Pub_library/Glossary.htm)

- We might find:
 - Explicit and implicit logic models
 - More than one logic model
 - Intended and unintended outcomes

Example: Intensive family intervention to reduce ASB



Inputs includes the resources used by projects which might be staffing, voluntary workers etc

Processes might include project planning, auditing, deploying resources etc

Context might include the baseline position (eg level of ASB, teenage pregnancy, school achievement) etc

Outputs might include number of families engaged, number of ASB contracts completed, number of case worker contacts etc

1st order outcomes might be reduction in anti-social behaviour etc

2nd order outcomes might be reduction in fear of crime, improvements in educational attainment, health improvements

Unintended outcomes might include (negative) displacement of ASB or (positive) diffusion of benefits etc

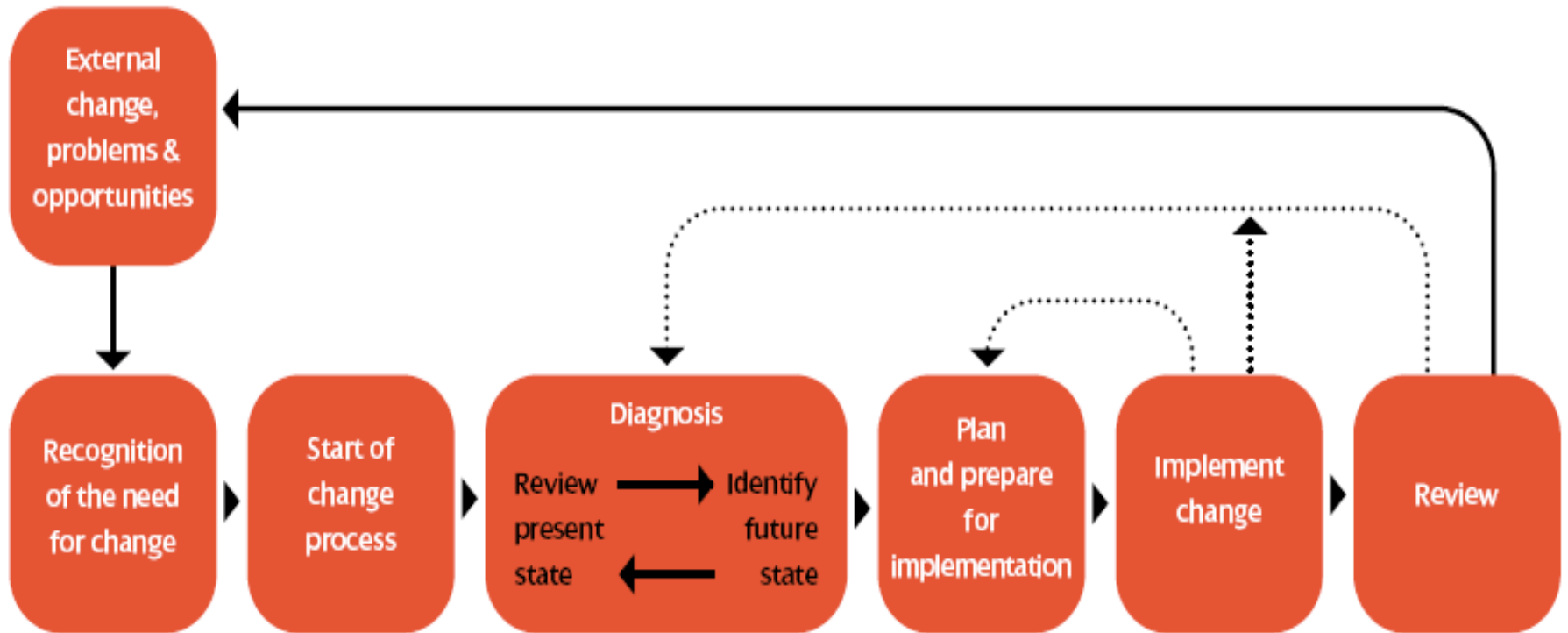
Can it work? Process evaluation

- Identify whether the intervention has been implemented as intended
- Explore how a service or policy is delivered and experienced in practice
- Identify the mechanisms by which it can produce the desired effects, potential barriers and facilitators
- Identify circumstances under which successful operation might be replicated

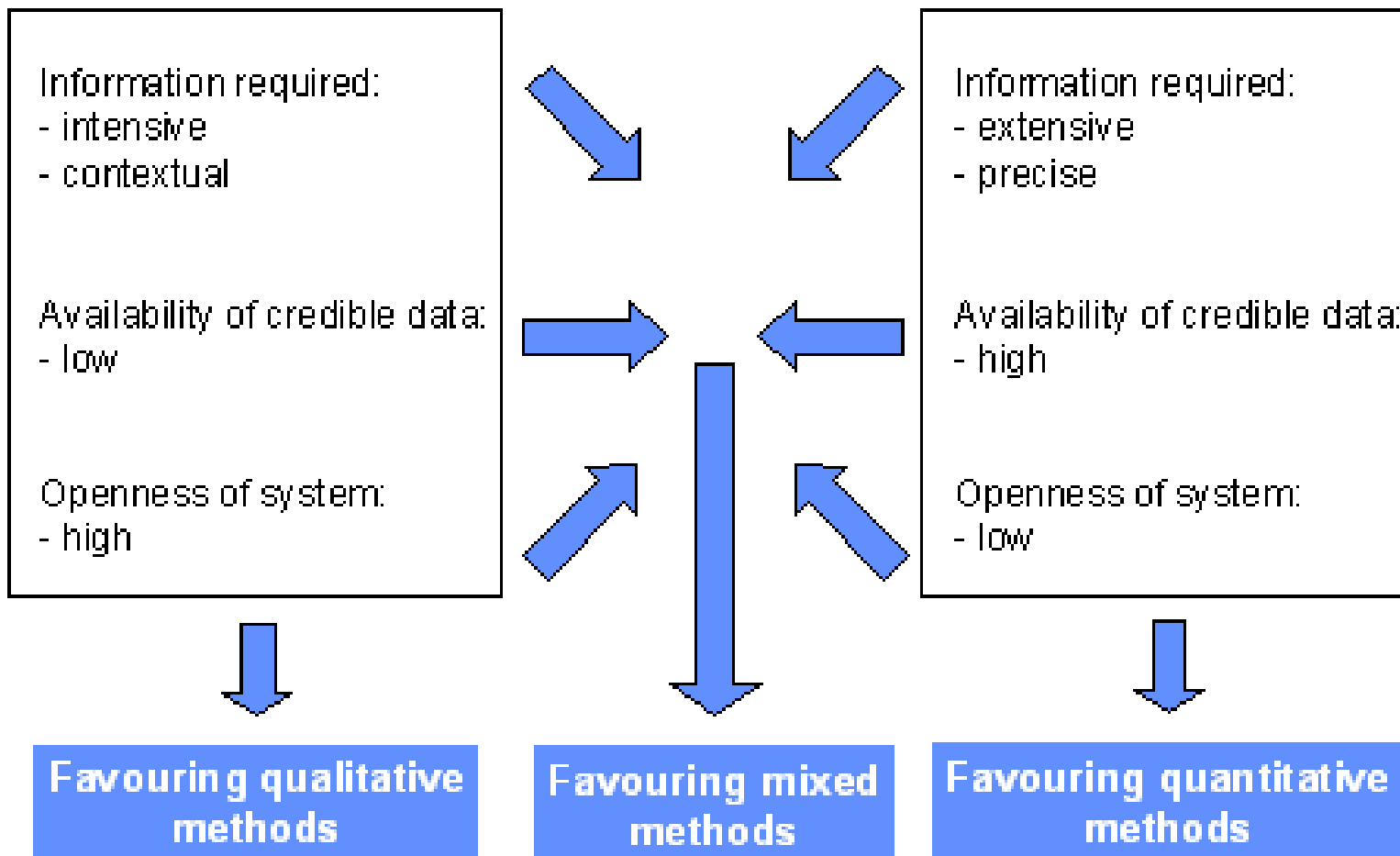
Key evaluation challenge

- A process evaluation will throw up a range of competing accounts of how a project has been implemented, what worked and what didn't work.
- How can we develop a framework for a process evaluation that allows us to make sense of these accounts?
 1. Draw on OD literature and management consultancy tools.
 2. Triangulate the evidence.

Draw on OD literature eg understanding change process



Triangulate evidence



Based on Chen, H. (1998), Reconciling conflicts and expanding options in policy or program evaluation. In Evaluation des politiques publiques, Kessler Mari-Christine, Lascoumes Pierre, Setbon Michel, and Theonig Jean-Claude (eds.), Paris, L'Harmattan.

Does it work? Impact evaluation

Scale	Impact evaluation methodology	Necessary conditions
5	Random Control Trial: Random assignment and analysis of comparable units to program and comparison groups.	Ability to randomly assign participants/non-participants to projects
4	Matched pairs: A comparison between multiple units with and without the intervention; or using comparison units that evidence only minor differences.	Close match between group of programme participants and non- participants
3	Multivariate model: A comparison between two or more comparable units of analysis, one with and one without the intervention, where there are differences in the relevant characteristics of the units.	Existence of comparable group of non-participations
2	Before/after or time series analysis: temporal sequence between the intervention and the measure.	Ability to measure before and after intervention
1	Correlation: Correlation between an intervention and a measure at a single pointing time.	Availability of data to show correlation

Sherman et al. (1998) *Preventing Crime. What works, what doesn't, what's promising.*
National Institute of Justice.

Measuring impact

- Choice of impact measures is often limited by availability of data or resources to collect primary data.
 - In the UK, re-offending data can be accessed via the Police National Computer (PNC). However,
 - Data is not always up-to-date and accurate
 - Access can be challenging even for a state-sponsored evaluation
- For less tangible outcomes such as family reconciliation or increased self-confidence the only option is often expensive face-to-face interviews/surveys.

Finding a comparator

- A comparator group or area may not be readily available.
- While a programme or pilot area may be willing to cooperate with an evaluation an organisation that is not part of the evaluation but is being asked to provide data for a comparator area may be hard to persuade to cooperate.
- A regional or national roll-out may limit the number of potential comparator areas
- Other interventions?

Throughputs

- Programmes and projects often struggle to generate sufficient throughputs to allow for robust statistical analysis.
 - Set-up times are often under-estimated
 - The need for inter-agency cooperation is often under-estimated

Timescales

- Often evaluation timescales preclude appropriate follow-up periods to identify the long-term impact of interventions.
 - The Home Office recommends that reconviction studies use a 2 year follow-up period.
 - By the time a cohort to study has been generated this will often require an evaluation period of 3 – 4 years.

Systematic Reviews

- Systematic reviews are overviews of the existing research literature on a topic based upon comprehensive searching of print, electronic, and unpublished sources. They provide a means of distinguishing the quality of research studies and presenting the available evidence in a way that makes the quality of the evidence upon which the review is based transparent.
- The Magenta Book suggests that a systematic review should address a question that has the following four components:
 - A clear specification of the interventions, factors or processes in question
 - A clear specification of the population and/or sub-groups in question
 - A clear specification of the outcomes that are of interest to the user of the review
 - A clear specification of the contexts in which the question is set.

Government Social Research Unit (undated) *The Magenta Book* www.policyhub.gov.uk/magenta_book

Rapid Evidence Assessments

- A tool for identifying and summarising available research evidence on a policy issue, as comprehensively as possible, within time constraints.
- Can be commissioned to specialists for relatively low cost.
- Key elements:
 - Agreed search strategy rigorously followed (eg defined search terms, list of databases and publications that will be searched)
 - Criteria set for data quality and only data that meets criteria is included in the synthesis
 - A REA report will provide overview of what evidence is saying. Sometimes data synthesis is undertaken if data being reviewed is quantitative and sufficient studies of an agreed methodological quality exist to make such an approach possible.

Cost Benefit Analysis (CBA)

- Three pieces of information required

The extra outcome achieved by the intervention compared with an alternative interventions

Derived from either an empirical impact study or a REA

The economic value of these outcomes

In the UK there is extensive data on the values of different outcomes

The extra cost of implementing the intervention compared with an alternative intervention

Derived from empirical investigation