



Right method, right price: the economic value and associated risks of experimentation

Gloria Laycock¹ · Jacqueline Mallender²

Published online: 9 October 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract

Objectives To consider the important contribution to crime reduction policy and practice made by research in experimental criminology; possible future research priorities and their methodological implications bearing in mind the cost and associated risks of research.

Method Discussion of the concepts necessary to inform policy and practice leads to a consideration of the appropriate methodology for primary research. On the basis of this discussion, three case studies are considered as exemplars of the argument being developed.

Results The authors argue that experimental criminology has been important in demonstrating the impact of certain types of place-based or people-based interventions. Nevertheless, the promised economic benefits are often predicated on interventions being proven effective in a particular geography or environment, on pre-existing levels of investment in the criminal justice system and on the presumed mechanism through which the initiative achieves its effect. As such, these ‘confounding factors’ need to be well researched and reported at the level of an individual experiment.

Conclusions Experimental criminology has made an important contribution to policy and practice, but could increase its relevance and impact by adopting evaluation methods which expose the risks of getting the wrong answer and demonstrate the extent to which risks, and therefore costs, might be reduced through strong study design and reporting.

Keywords Experimental criminology · Evidence-based policy · Randomized controlled trials

✉ Gloria Laycock
g.laycock@ucl.ac.uk

Jacqueline Mallender
jacque.mallender@optimityadvisors.com

¹ UCL Jill Dando Institute of Security and Crime Science, London, UK

² Partner Optimity Matrix, London, UK

Introduction

There is a growing body of literature using experimentation to demonstrate the potential effectiveness of crime reduction policies, much of it published in journals such as *The Journal of Experimental Criminology*. Moreover, the expected economic benefits of these policies to taxpayers, victims and wider society can potentially exceed costs several times over. It is perhaps timely, therefore, given the 10th anniversary of the Journal, and the current interest in evidence-based policy and practice, to review the wider contribution of experimental criminology to the determination of what works and at what cost – have we got it right?

Public policy researchers regularly engage in healthy debate about the role of experiments and how these compare with alternative research methods. Criminologists are no exception. What is in no doubt is that experimentation in criminology, and indeed wider social policy, is here to stay. In the USA, the Office of Management and Budget (OMB) guidance for budget submissions in 2015 placed the onus on agencies to demonstrate evidence for spending programmes through “high quality experimental and quasi-experimental studies” (OMB 2013). In the UK, a recent paper from the Cabinet Office promotes the wider use of randomised controlled trials (RCTs) (Haynes et al. 2012) across Government and is now accompanied by a nascent Cross-Whitehall Trial Advice Panel for Departments wishing to design RCTs (Cabinet Office 2014). Whilst it is still considered by its advocates to be an under-used methodology, the increasing adoption of it as a ‘gold standard’ by policy makers will undoubtedly redress this perceived imbalance.

Against this background, the present paper is specifically concerned with the design of experimental primary research in criminology. Section 2 considers the important role now played by experiments in criminology and the promised economic benefits. Section 3 analyses the issues that impact on the translation of research into practice and the implications for research methods. It also presents a discussion on the growing use of decision modelling to measure economic benefit. Section 4 considers three case studies. The final section draws some conclusions and makes recommendations for the way in which evaluative research might be carried out in future.

The important role of experimental criminology

There are perhaps three reasons for the current interest and support for experimental criminology: First, experimental methods in medicine are seen as having delivered huge benefits in terms of reduced mortality, increased wellbeing and reduced costs, and it is hoped that by encouraging similar approaches in the field of criminal justice commensurate gains might be made. Secondly, there is pressure on the public sector across the board to deliver more for less – determining ‘what works’ in the delivery of effective public services, including those associated with policing and criminal justice more broadly are therefore ripe for development. Finally, if you do not include evidence as part of the decision process what are you left with? Public policy in general and criminal justice delivery in particular are not art forms: They can and should be informed by established knowledge. That is not to say that policies should not also be

influenced by other issues such as ethics, justice, politics, precedent and so on, but it is to say that evidence can and should play a part and arguably a greater part than hitherto.

It is against this background that experimental criminology is developing. Experimental criminology covers research designed to measure the impact of an intervention to reduce crime and fear of crime using well-designed RCTs or quasi-experimental studies. Proponents of experimental methods argue that these approaches reduce the risk of bias in results. Such bias occurs when the measured outcome of an intervention is affected by confounding factors, particularly those which are unknown in advance and/or cannot be measured or controlled for. There are many good discussions on what defines a good quality experiment. In October 2014, the Coalition for Evidence-Based Policy published a useful overview (CEBP 2014). This paper sets RCTs as the strongest method and cites many authoritative scientific bodies by way of support, including the National Academy of Sciences (NRCIM 2009), the National Science Foundation (NSF 2013), the Institute of Education Sciences (IES 2013) and the Food and Drug Administration (FDA 2008). The UK Government's Behavioural Insight team also supports RCTs as providing the best method for evaluating impact, although they are still relatively under-used in public policy (Haynes et al. 2012). Hough (2010), on the other hand, argues that the use of experiments leads to bias in research and results in an over-investment in treatment-effect (technocratic) models at the expense of models that explore the more complex role of behaviour change in offenders. The debate will undoubtedly continue and, whilst interesting and important to the future of criminology research generally, it is not the focus of this paper.

However designed, experiments in social policy are playing an increasing role in demonstrating and measuring the impact of certain types of place-based or people-based interventions. Much of this work is now synthesized in the form of international systematic reviews and published by organisations such as the Cochrane Collaboration, the International Initiative for Impact Evaluation (3ie), and the Campbell Collaboration.¹

The Campbell Collaboration was established in 1999 and, at the time of writing, has published over 100 systematic reviews. The Campbell Collaboration is a sister organisation to the Cochrane Collaboration, which was established in 1994 and draws on a network of independent practitioners and researchers to undertake and publish systematic reviews in human healthcare and health policy. Now covering centres in around 120 countries, there are over 5000 Cochrane Reviews, and the Cochrane Database of Systematic Reviews carries an Impact Factor for the CDSR of 5.785 in 2012 and is ranked 11th of 151 journals in the "Medicine, General & Internal" category of the Journal Citation Reports[®] (JCR).²

Both Collaborations take experimentation as a baseline for assessing the quality of primary research on effectiveness, with RCTs as the 'gold standard', although partly in response to criticism (see Greenhalgh et al. 2014 for an argument in relation to healthcare) other forms of evidence are increasingly being included in reviews. The number of Cochrane reviews (5000) relative to Campbell reviews (100), however,

¹ See, for example, the reviews published on the Campbell Collaboration website. www.campbellcollaboration.com

² <http://www.cochrane.org>

partially reflects the relative differences in use of high quality RCTs as the standard method for assessing the effectiveness of clinical treatments for individuals amongst practitioners, industry and governments. Whilst RCTs are now considered the norm in health care, to a considerable extent funded by the drugs companies with their own imperatives, the same is not true in social policy (Haynes et al. 2012).

The Crime and Justice Coordinating Group is one of five topic groups in the Campbell Collaboration, and, at the time of writing, has published 38 systematic reviews, with a further 35 reviews at various stages of development. The published reviews cover interventions targeted at the individual, place-based interventions, and institution-based interventions. The reviews are intended to provide information for policy makers on the effectiveness of interventions designed to prevent crime, correct behaviours with a view to reducing reoffending, or to impact on the wider determinants of crime for individuals or communities.

In the USA, such evidence forms part of a broader evidence base for policy makers at national or state level through centres such the Institute of Public Policy in Washington State (WSIPP)³ and other similar state-level initiatives supported by the Pew-MacArthur Results First Initiative,⁴ or through organisations such as The Centre for Problem-Oriented Policing⁵ and the National Institute of Justice, Crime Solutions.⁶ What is interesting about these centres is that they seek to provide a context specific window into the evidence such that policy makers can see what is already known about an intervention from research elsewhere in the international, national, state or local contexts.

The WSIPP model is particularly interesting in that it looks at interventions and their relative costs and benefits in the specific context of Washington State. Essentially, WSIPP compiles potential effect sizes for interventions based on systematic reviews of experimental crime research (including Campbell reviews), calculates effects on a consistent basis across interventions, and estimates their relative costs and benefits from the perspective of the Washington State tax dollar and the wider societal economic impact. It is estimated that using this tool to prioritise programmes has resulted in savings of \$1.3 billion over a 2-year budget cycle⁷ (Urahn, 2012). The Results First Initiative is seeking to replicate and develop this further in other states across the USA.

Another example of the growing use of experimental evidence in social policy can be seen in the UK. Initially established in 1999 to work in healthcare, the National Institute of Health and Care Excellence (NICE) now covers public health (since 2005) and social care (since 2013). One aspect of NICE is to provide guidance to the NHS on policy and practice. The guidance draws on systematic reviews of evidence, combined with cost-benefit analysis, to assess the value of the adoption of health and social care interventions in the context of the NHS. In 2013, the UK Government launched the What Works Network in social policy,⁸ to include NICE but expanded to cover six new centres, including:

³ <http://www.wsipp.wa.gov/>

⁴ <http://www.pewtrusts.org/en/projects/pew-macArthur-results-first-initiative>

⁵ <http://www.popcenter.org/>

⁶ <http://www.crimesolutions.gov/>

⁷ <http://www.governing.com/columns/mgmt-insights/col-cost-benefit-outcomes-states-results-first.html>

⁸ <https://www.gov.uk/what-works-network>

- Sutton Trust/Educational Endowment,
- What Works Centre for Wellbeing,
- Early Intervention Foundation,
- What Works Centre for Local Economic Growth,
- Centre for Ageing Better, and
- What Works Centre for Crime Reduction.

Of particular relevance to this paper is the What Works Centre for Crime Reduction (WWCCR)⁹ which has been set up to support crime reduction practitioners, including the police and Police and Crime Commissioners (PCCs) by synthesising evidence on interventions and operational practices which reduce crime and providing associated tools and guidance to help with local prioritisation and implementation decisions. The centres all take a relatively inclusive view of evidence, complementing experimental evidence with other high quality study designs and expert elicitation.

One of the drivers for the establishment of these centres is the growing acceptance that crime reduction programmes and policies to reduce reoffending, if effective, are generally cost beneficial (Mallender and Tierney 2013). As mentioned above, the economic modelling work undertaken for WSIPP suggests that:

- For juvenile justice, 10 of the 11 interventions covered by the WSIPP systematic reviews show positive economic benefits when subject to economic analysis. Estimates of benefit–cost ratios relevant to Washington State range from 2.51 to 41.75.
- For adult justice, 16 of the 18 interventions covered by the WSIPP systematic reviews provide positive benefit–cost ratios ranging from 1.93 to 40.76.
- For policing, the analysis shows a \$6.94 benefit–cost ratio for the adoption of deployment of one additional police officer on hot-spots policing within Washington State.

Taken at face value, these are impressive figures. According to a recent report from the Results First Initiative, applying the model to five other states (Iowa, Massachusetts, New Mexico, New York, and Vermont) has resulted in a realignment of \$81 m towards effective programmes that work (Pew 2014).

The prize is substantial. If research can demonstrate which programmes and policies are both effective and cost efficient, and policy makers and practitioners can implement these successfully, this will result in the double benefit of reducing crime and reoffending and saving money.

Real world considerations in translating policy to practice

In practice, the promised benefits are difficult to achieve. This may be due to prevailing local circumstances, the adequacy of investment in new programmes, the political difficulty of withdrawing from much loved programmes that do not work (or do not work well), and the relative efficacy of implementation. It may also be due to the very

⁹ <http://www.college.police.uk/en/20399.htm>

real difficulties of translating findings generated from experimental research from one context to another.

Context matters: mechanisms, moderators and implementation issues

A consortium of eight universities¹⁰ supports the UK What Works Centre for Crime Reduction. The task of the consortium is, *inter alia*, to identify, retrieve and summarise the results of all existing systematic reviews or meta-analyses on what works in crime reduction; to carry out 12 new reviews over the 3 years of the project; and, importantly, to provide the results of this exercise in a format that can be used by policy makers and practitioners with a remit to reduce crime – largely but by no means exclusively the police.

The view taken by the research consortium is that policy makers and more so practitioners need information under five headings in order better to inform decision making about the adoption or otherwise of a given initiative. They need to know the effect size (E) of the intervention; how this effect is achieved, that is, the mechanism (M); what moderates (M) this effect, that is, the local context; what is known about implementation issues (I); and what can be said about costs and other economic considerations (E). The acronym EMMIE spells out these main dimensions. Full details on the definition of these concepts and discussion on the detail of EMMIE can be found in Johnson et al. (2015).

In practice, however, it has become clear in identifying existing systematic reviews, and from the first tranche of the 12 new reviews, that the primary research literature has little if anything to say on the key elements of mechanism, moderator and implementation (MMI), and such information as is available on costs is sparse. There is also a need to link these dimensions together in a more systematic manner. For example, what might be cost effective in one jurisdiction or local area may be an economic disaster in another.

Pawson and Tilley (1997) give a full description of mechanisms and contexts as they relate to crime and criminology. By mechanisms they refer to *how* an initiative might take effect. In addition, whether or not any particular mechanism is fired will depend on the context within which it is introduced. What works in the USA may not work in England; what works in London may not work in Birmingham, and so on. In order to hypothesise (and note that it is a hypothesis) whether or not an initiative judged successful in one place and time will work in another place and time the practitioner (or policy maker) will need to understand the mechanism and its relationship with the context.

Sidebottom and Tilley (2011), picking up on these issues, have reviewed reporting guidelines in health and criminology, and in particular discuss the recommendations made in an article by Perry et al. (2010) which reviewed the extent to which studies using RCTs had used existing reporting guidelines. The general conclusion was that these studies were in the main poorly reported. But Tilley and Sidebottom go further in arguing for more systematic reporting in relation not only to research using RCTs, but also to criminological research more broadly. In particular, and notwithstanding the

¹⁰ University College London (UCL), the Institute of Education (IoE), the London School of Hygiene and Tropical Medicine, Birkbeck College, and Cardiff, Dundee, Surrey and Southampton universities.

methodological inclinations of researchers, they argue for the addition of two items: “[T]he first relates to the causal mechanisms with which an intervention is believed to operate. The second relates to the setting, circumstances, and procedures of implementing an intervention” (Sidebottom and Tilley 2011:54). These items are not only important in their own right, but are also proving crucial in contributing to improvements in systematic reviews and meta-analyses, as the UK What Works exercise is proving.

Sampson et al. (2013) reinforce this point in discussing recent attempts by criminologists to increase the influence of their research on policy. Like Sidebottom and Tilley, they note the importance of context to the translation of research into practice. They also implicitly criticise the research methods used in many RCTs which admirably demonstrate internal validity but fail abysmally when considering the external validity – which is precisely what is needed if research is to be generalised and therefore of use to policy makers and practitioners. They say:

Even with no effect heterogeneity and full knowledge of the mechanisms operating within a particular study, the context challenge implies that a single experiment cannot provide evidence of the consequences of scaling up. Although this challenge might not matter in medical trials, the canonical example of an experimental science, crime, and criminal justice are quintessentially social phenomena (Sampson et al. 2013:20).

A rethink is required which acknowledges the need to develop middle-range theories and which takes account of the need to balance internal and external validity with greater understanding of the change mechanism, all delivered at reasonable cost and in timely fashion.

If mechanisms and contexts have been relatively neglected by researchers, this is even more the case in relation to the detail of implementation. Rosenbaum (1986) made the point that there are three potential sources of failure in crime reduction – theory failure, measurement failure and implementation failure – and all three need to be guarded against in the course of primary research. Research evaluations routinely test for theory failure – this is in essence a search for effect. Measurement failure is less commonly discussed but is often considered in the field of crime reduction and may, for example, point to the need for crime surveys rather than reliance upon police crime statistics. Although it is now commonplace to guard against implementation failure by ensuring, for example, that measures are implemented as planned, there remain many elements of implementation that are not considered. For example, evaluations of the effectiveness of CCTV rarely consider or discuss the type of CCTV system being introduced, the characteristics of the cameras, the direction in which they may or may not be pointing, how many (if any) staff are routinely watching screens, how well advertised the presence of the system may or may not be, and so on. These factors are inevitably relevant to both the likely efficacy of any system (the benefits) and, importantly, to its associated costs.

Thinking through mechanisms, contexts and implementation issues is tricky, but it is important because from a research perspective it affects the choice of methodology and, importantly for this article, the cost of the research itself. It also severely limits the

ability of those attempting to develop the middle range theories needed by practitioners through a research synthesis process.

Implications for research methods

The medical research literature has proved enormously influential in promoting the adoption of the RCT (plus double-blind frills when possible) as the ‘gold standard’ for experimental research in criminology (Sherman et al. 1998). Over recent decades, however, even medical research has become much more sensitive to the mechanisms through which an intervention might exert its effect, and to the context within which it might do so. As medical treatments move from the distribution of pills dispensed to tackle well-understood medical conditions to personalised therapies designed to target particular malfunctions in DNA structure (to take an extreme example), RCTs have become more complex to design and implement and average effect size is no longer the only measure of interest. The importance of clinical judgement and relevant context, and their intimate relationships with outcome is increasingly being recognised in the medical literature (Greenhalgh et al. 2014). If this is true for medical research, it is even truer for social science.

Cartwright and Hardie (2012) support the view that more information is needed for experiments to play their full role in policy development. In the preface to their recent book, they say:

You are told: use policies that work. And you are told: RCTs – randomized controlled trials – will show you what these are. That’s not so. RCTs are great, but they do not do that for you. They cannot support the expectation that a policy will work for you. What they tell you is true – that this policy produced that result there. But they do not tell you why that is relevant to what you need to bet on getting the result you want here. For that, you will need to know a lot more (Preface).

In considering the role of RCTs in International Development, White makes the point that “So-called ‘black box’ impact evaluations which do not seek to unpack the causal chain, to understand why a programme does or does not work in a particular setting, are of far less benefit to policy makers than those that do” (White 2013).

In other words, although, where applicable, RCTs and the systematic reviews of RCTs directed at assessing the efficacy of social policy are good at minimising the risk of attribution bias (and thus ensuring internal validity), they are likely to be of more use to a wider audience if they are part of a wider research programme. This wider perspective would also look at and report on mechanism, context and implementation, and, in so doing, report on resource use and other factors likely to impact on cost.

Economic decision models

Policy makers use economic evaluation to identify the resources needed to fund a policy intervention, to identify the associated return on investment for taxpayers and the wider society, and to assess the affordability of a package of interventions or a

programme given competing priorities for funds. Cost–benefit analysis has become a useful analytical framework for answering these questions (Cohen 2000).

The economics of the policy is hugely impacted by mechanism, context and implementation factors. For these reasons, economists generally use context-specific decision models to identify the costs and benefits of policy interventions. As defined by Philips et al. (2004), decision analysis is a structured way of thinking about how an action taken in a current decision environment would lead to a result. It usually involves the construction of a logic model, which is a mathematical representation of the relationships between inputs and outputs and outcomes in the context of the proposed adoption of a policy and/or the implementation of an intervention. Decision models form the basis of the economic modelling used for the WSIPP and Results First Initiative (Aos et al. 2011).

Whilst there is an increasingly rich source of good quality economic models undertaken as part of primary research studies, there are strong arguments as to why systematic reviews of economic evaluations are of limited value (Anderson and Shemilt 2010). These arguments centre on the importance of context, which itself impacts on the choice of model, the scope of the economic analysis, and the methods used to calculate costs and benefits. There is a role for systematic reviews of economic studies to help to map theories, and to help provide data and information to populate new economic models and to source relevant studies for a specific topic or intervention. However, demonstrating the economic value of a policy or intervention requires that it be assessed in context (Mallender and Tierney 2013).

So what does “in context” mean in these circumstances? To understand this, it is worth looking more closely at how decision models are constructed. Most commonly, the objective of decision models is to understand the relationship between incremental cost and effect in order to assess relative cost effectiveness, and to determine which interventions should be adopted given existing information. Typically, a decision model relies on:

- the cost of the resources required to deliver the intervention,
- the effect of the intervention on the outcome of interest, and
- the economic value associated with the health outcome generated by the intervention – usually expressed in terms of quality of life gains and health care cost savings.

A simple graphical representation of a decision model is shown in Fig. 1. The costs and benefits are calculated by comparing the difference between the two decision arms (new programme compared with the alternative) also taking into account temporal issues and intangible costs and benefits. Models can be constructed as simple static deterministic models, or dynamic stochastic models using methods such as Monte Carlo simulation, to estimate long-term costs and benefits.

Several guidelines have been published for those developing and evaluating decision-analytic models. One of the most comprehensive has been developed by NICE in England (NICE 2012). In addition to describing how to use reviews of economic studies and the development of economic models to inform guidelines, the guidance prescribes a mixed method approach to gathering evidence. Essentially, these models need both scientific and other types of evidence from “multiple sources,

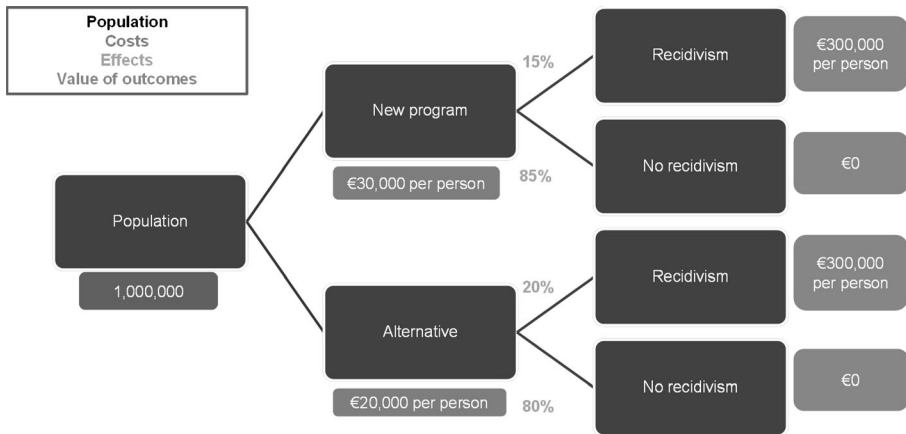


Fig. 1 Graphical Representation of Economic Decision Model

extracted for different purposes and through different methods... within an ethical and theoretical framework". Evidence is classified into either *scientific evidence*: which is defined as "explicit (codified and propositional), systemic (uses transparent and explicit methods for codifying), and replicable (using the same methods with the same samples will lead to the same results); it can be context-free (applicable generally) or context-sensitive (driven by geography, time and situation)", or *colloquial evidence*: essentially derived from expert testimony, stakeholder opinion and necessarily value-driven and subjective.

For the economic decision models themselves, issues focus on:

- structure, including for example, statement of the decision model, model type, interventions and comparators, and time horizon;
- data, including data sources and associated hierarchies (Coyle and Lee 2002), data identification, data modelling, and assessment of uncertainties; and
- consistency, including internal and external consistency or validity of the model.

Increasingly, governments are supporting local agencies to use this approach to undertake cost–benefit analysis in a local context and are providing data to support them. For example, knowledge banks are being developed such as the Cost-Benefit Knowledge Bank for Criminal Justice in the USA (CBKB 2013) and the cost of crime statistics published by the UK Home Office in 2011 (an update of Dubourg et al. 2005).

Factors likely to impact on costs and benefits which will vary according to context include:

- prevailing target population characteristics and the local 'epidemiology' of crime and recidivism – this will include age of offenders, prevailing age–crime cohorts, availability of employment, housing and healthcare, and many other factors;
- prevailing crime and corrections investment including number of practitioners, pre-existing skills and capabilities, existing operational practices, and the age and quality of the criminal justice estate;

- the implementation requirements of the programme and, specifically, the processes required to ensure programme efficacy— this will impact on the likely marginal requirements for investment in people, process and infrastructure;
- the existence of other programmes in the local area which might differ from the comparator used for the RCT and would alter the predicted effect size; and
- prevailing budgetary systems which could determine whose budget is going to be impacted by changes in costs and benefits and hence whose perspective will be important in driving a decision to invest.

The decision itself will also vary depending on whether the analysis is being done at local, regional and/or national level.

For decision makers to assess ‘what will work, here, now and in this place’, they will need information about the intervention which will allow them to adjust and adapt the parameters of the decision model to suit their local circumstances. Essentially, in addition to any effectiveness measure, they will need information about context, mechanism, moderators and implementation.

Three case studies

We have selected three case studies to illustrate the very real difficulties policy makers currently have with identifying experimental research and using this research evidence to inform practice. The first deals with custodial versus non-custodial sentencing and what we can learn from experimental studies. The second looks at offender-based programmes and what the evidence says about cost benefit. The third case study looks at a specific place-based intervention, namely Neighbourhood Watch, and what research can tell us about mechanisms, moderators, and implementation issues.

As has been shown by the WSIPP experience, policy makers are increasingly interested in the efficacy of prison. Prison is an expensive resource. In the USA, state prison costs \$31,206 per prisoner per year. Whilst incarcerated, the risk of recidivism is put on hold (if we disregard offending in custody); however, post-release, is there any evidence that prison has impacted on reoffending as compared with community supervision?

The most recent systematic review of prison versus community corrections is the Campbell Review by Villettaz et al. (2015). The authors reported the results of an extensive systematic review of experimental or quasi-experimental evaluations of the effects of custodial and non-custodial sanctions on reoffending. This was an update of an earlier review conducted in 2006. In total, the authors examined more than 3000 abstracts, and identified four RCTs, two natural experiments and ten matched-pair design studies or quasi-experimental studies (QES) which met the criteria set out in the review protocol. The authors undertook a separate meta-analysis of the RCTs and the QES and found that the RCTs showed no differences in reoffending, whilst the QES suggested that non-custodial sentences were more effective. The authors conclude that the use of QES may not correct sufficiently for differences between the comparison and the control group and that “the problem is that comparisons between custodial and non-custodial sanctions are systematically biased because, as Bales and Piqueiro (2012) phrased it, “the main problem in this area of research is that individuals sentenced to

prison differ in fundamental ways from those individuals who receive a non-custodial sanction” (p. 97).

Looking closely at the studies included in the meta-analysis, it is possible to see that both the RCTs and QES are highly disparate in terms of the target populations (juveniles vs. adults), the programmes that accompanied the interventions (training, supervision, no support), and the length of follow-up. Whilst this heterogeneity is not in itself a problem, the results of the meta-analysis and even of the systematic review itself show that existing evidence from experimental studies is potentially of limited use to policy makers trying to grapple with the complex issue of sentencing policy. What this research certainly tells us is that there remains a dearth of experimental studies comparing custodial and non-custodial sentencing. The authors could not find evidence that a single RCT had been undertaken since the review in 2006.

Moving to the second case study, the adult offender programmes within the WSIPP model identifies five programmes generating the best economic return (benefit–cost ratio). These are shown in Table 1. By comparison, also in the table are the programmes ranked in the top five in terms of savings to Washington State taxpayers. Interestingly, three interventions are good for taxpayers and the wider economy: electronic monitoring, correctional education in prisons, and vocational education in prison.

A similar analysis has been undertaken in the context of the UK (Tierney and Mallender 2015) as shown in Table 2. The work represents an update of *The Economic Case For and Against Prison* (Matrix Knowledge 2008). This analysis shows a different top five in terms of savings to taxpayers and economic return on investment. Only two interventions appear on both lists: residential drug treatment and surveillance + drug treatment.

Given that both use relatively similar approaches to identifying evidence, and modelling economic return, what are the main reasons for the differences? In summary, the key differences appear to be:

- the methods of extrapolating the impact over time – most of the primary study data examine impact 1–2 years post-intervention; the WSIPP model and the Matrix model draw from different sources to determine age–crime curves and the longevity of the effect;
- the actual timeframe of analysis;
- the counterfactual used to model the impact of the intervention and the costs and benefits; and

Table 1 Comparison of WSIPP top 5 rankings on two measures

WSIPP (adult crim. just.): Top 5 Savings to Taxpayers	WSIPP (adult crim. just.): Top 5 on Economic Return
<ul style="list-style-type: none"> • Electronic monitoring • Correctional education in prison • Vocational education in prison • Outpatient/non-intensive drug treatment in prison • Drug Offender Sentencing Alternative 	<ul style="list-style-type: none"> • Offender Re-entry Community Safety Program • Therapeutic communities for offenders with co-occurring disorders • Electronic monitoring • Correctional education in prison • Vocational education in prison

Table 2 Comparison from Tierney and Mallender (2015) top 5 rankings on two measures

Tierney and Mallender (adult crim. just.): Top 5 Savings to Taxpayers	Tierney and Mallender (adult crim. just.): Top 5 on Economic Return
<ul style="list-style-type: none"> • Residential drug treatment • Prison+behavioural interventions for juveniles • Surveillance+drug treatment • Community service+behavioural interventions • Community service 	<ul style="list-style-type: none"> • Residential drug treatment • Prison+drug treatment • Prison+educational/vocational interventions • Surveillance+drug treatment • Boot camp for juveniles

- methods for estimating intangible and wider societal benefits (such as costs incurred in anticipation of crime, e.g. security expenditure), as a consequence of crime (such as property stolen and emotional or physical impacts), and in response to crime (costs to the criminal justice system), and the lost economic output of offenders.

In other words, one has been developed for use in Washington State and the other for use in the UK.

Going one stage further, do either of the models provide sufficient information sourced from the primary studies or systematic reviews for policy makers to decide whether and how to implement any of these programmes nationally or locally? The answer is almost certainly ‘no’. Neither model provides sufficient information about why these programmes have worked in their original setting, in what context, and what might moderate their impact. Both would benefit from a more comprehensive and systematic reporting of mechanisms, moderators, and implementation issues.

Moving on to the third and final case study, the relevant Campbell Review is the “Effectiveness of Neighbourhood Watch: A Systematic Review” (Bennett et al. 2008). The review authors undertook a systematic review of neighbourhood watch schemes to assess their potential impact on crimes. The review covered schemes that were either stand-alone, or included additional programme elements such as property marking and security surveys. The high level finding of the review is that “neighbourhood watch is associated with a reduction in crime”. However, a deeper look at the review demonstrates a much more mixed evidence base and a paucity of well-conducted experiments, primarily because of the difficulties of engaging neighbourhoods who do not want to participate.

The authors provide a very useful description of the history and theory behind neighbourhood watch; they also discuss the programme elements (mechanisms) and the likely moderators. The synthesis comprises a narrative review (19 studies covering 43 evaluations) and a meta-analysis (12 studies covering 19 evaluations) and includes a detailed synopsis of each study. However, despite best efforts, there were insufficient data reported in the original primary studies to enable the authors to identify why neighbourhood watch is associated with a reduction in crime. Whilst the authors had proposed a theory base to the intervention, the studies did not provide information on mechanisms, and the authors reported that it was difficult to say why these schemes might work.

This difficulty makes it even harder for policy makers and/or local agencies to assess whether such schemes might be relevant and impactful in their local areas. In this example, the review authors were keen to identify mechanisms, moderators and implementation issues and would have reported these; however, the primary studies

did not include sufficient information to provide a real understanding of why this policy might work 'here and in this place'.

Conclusions

This paper has demonstrated the necessity for primary researchers, systematic review authors, and policy advisors to take a broader look at evidence. Standards of evidence have been improving and the appropriate use of experimentation to mitigate the risk of bias is a welcome addition to the portfolio of methods available. However, simply reporting on the effectiveness of a programme is of limited value to policy makers and practitioners. The case studies have shown that high-quality analysis on the part of systematic reviewers and modellers is hampered by primary research, which is narrow in scope and often poorly reported. We would encourage all those dedicated to improving the quality of research to also look at improving the relevance of research to policy makers and local agencies in other jurisdictions who are looking to translate learning from one context to another (see also Laycock 2014).

We have referenced the EMMIE framework developed by the WWCCR as one potential tool that could be used by researchers and systematic reviewers to report their findings. In this way, policy makers and practitioners can interpret the evidence more effectively and can use research outputs to inform local jurisdiction-specific decisions. But the input to EMMIE is only as good as the research on which it draws. This primary research needs to go beyond assessments of effect size and take greater account of mechanisms and contexts.

We should stress that we are not expecting any one study to come to a definitive conclusion on all elements of EMMIE, but rather we wish to make the point that, if researchers continue with business as usual and focus on effect size to the exclusion of other highly relevant elements of effective practice, then we never will be in a more informed position. We are looking for a contribution from primary research to the incremental improvement of evidence in policy and practice on the basis of a large number of well-specified and diverse studies.

We would also encourage primary researchers to record detailed information about implementation issues, resources and costs; however, the use of economic analysis to assess the costs and benefits needs to be undertaken in context and, whilst decision models are a very useful tool for policy analysis, they are rarely generalizable.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Anderson, R., & Shemilt, I. (2010). The role of economic perspectives and evidence in systematic review. In I. Shemilt, M. Mugford, L. Vale, K. Marsh, & C. Donaldson (Eds.), *Evidence-based decisions and economics: Health care, social welfare, education and criminal justice*. Oxford: Wiley-Blackwell.

- Aos, S., Lee, S., Drake, E., Pennucci, A., Klima, T., Miller, M., Anderson, L., Mayfield, J., & Burley, M. (2011). *Return on investment: Evidence-based options to improve statewide outcomes. Document No. 11-07-1201*. Olympia: Washington State Institute for Public Policy.
- Bales, W. D., & Piquero, A. (2012). Assessing the impact of imprisonment on recidivism. *Journal of Experimental Criminology*, 8, 71–101.
- Bennett, T., Holloway, K., & Farrington, David. (2008). *The effectiveness of neighbourhood watch*. doi: 10.4073/esr.2008.18.
- Cabinet Office. (2014) Call for external expertise to join cross-whitehall trial advice panel.
- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford: Oxford University Press. 196 pp.: 9780199841622.
- CBKB (2013). About CBKB [Cost-benefit knowledge bank for criminal justice]. Retrieved from <http://cbkb.org/about/>.
- Coalition for Evidence-Based Policy (CEBP). (2014). *Which study designs are capable of producing valid evidence about a program's effectiveness: a brief overview*.
- Cohen, M. A. (2000). Measuring the costs and benefits of crime and justice. In G. LaFree (Ed) *Measurement and analysis of crime and justice*. Washington, D.C.: National Institute of Justice, US Department of Justice
- Coyle, D., & Lee, K. M. (2002). Improving the evidence base of economic evaluations. In C. Donaldson, M. Mugford, & L. Vale (Eds.), *Evidence-based health economics*. London: BMJ Books.
- Dubourg, R., Hamed, J. & Thorns, J. (2005). *The economic and social costs of crime against individuals and households 2003/04*. London: Home Office online report 30/05.
- Greenhalgh, T., Howick, J., & Maskrey, N. (2014). Evidence based medicine: a movement in crisis? *BMJ*, 348, g3725. doi:10.1136/bmj.g3725. Published 13 June 2014.
- Haynes, L., Service, O., Goldacre, B., & Torgerson, D. (2012). *Test, learn, adapt: developing public policy with randomised controlled trials*.
- Hough, M. (2010). Gold standard or fool's gold? The pursuit of certainty in experimental criminology. *Criminology and Criminal Justice*. doi:10.1177/1748895809352597.
- Institute of Education Sciences (of the U.S. Department of Education) and National Science Foundation, *Common guidelines for education research and development*, August 2013, retrieved from: <http://www.nsf.gov/pubs/2013/nsf13126/nsf13126.pdf>.
- Johnson, S. D., Tilley, N., & Bowers, K. J. (2015). Introducing EMMIE: an evidence rating scale to encourage mixed-method crime prevention synthesis reviews. *Journal of Experimental Criminology*. doi:10.1007/s11292-015-9238-7.
- Laycock, G. (2014). Crime science and policing: lessons of translation. *Policing*, 8(4), 393–401. doi:10.1093/policing/pau028. first published online September 9, 2014.
- Mallender, J. & Tierney, R. (2013). (forthcoming). *Why pay for crime control: lessons from reviews of economic studies of crime and justice interventions*. In. Farrington, D. et al. *Lessons from systematic reviews in crime and justice* (Springer).
- Matrix Knowledge. (2008) *The economic case for and against prison*.
- National Research Council and Institute of Medicine. (2009). In M. E. O'Connell, T. Boat, & K. E. Warner (Eds.), *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington DC: National Academies Press. recommendation 12–4, p. 371, retrieved from: <http://coalition4evidence.org/1399-2/national-academy-of-sciences-report/>.
- National Institute for Health and Care Excellence (NICE). (2012). *Methods for the development of NICE public health guidance (third edition)*, retrieved from: <http://goo.gl/zPAqrV>.
- Office of Management and Budget (2013). *Memorandum to the heads of departments and agencies: Next Steps in the Evidence and Innovation Agenda*. Retrieved from: <http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-17.pdf>.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Perry, A. E., Weisburd, D., & Hewitt, C. (2010). Are criminologists describing randomized controlled trials in ways that allow us to assess them? Findings from a sample of crime and justice trials. *Journal of Experimental Criminology*, 6, 245–262
- Pew (2014): Pew Center on the States: Results First (Overview). Retrieved from <http://www.pewstates.org/projects/results-first-328069>.
- Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, Woolacott, Glanville N. (2004). *Review of guidelines for good practice in decision-analytic modelling in health technology assessment*. Health Technology Assessment; Vol. 8: No 36.
- Rosenbaum, D. P. (1986). *Community crime prevention: Does it work?* Beverly Hills: Sage.

- Sampson, R.J., Winship, C., & Knight, C. (2013) *Translating causal claims principles and strategies for policy-relevant criminology*. doi:10.1111/1745-9133.12027 *Criminology & Public Policy*, Volume 12 Issue 4.
- Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P., & Bushway, S. D. (1998). *Preventing Crime: What works, what doesn't, what's promising*. Washington, D.C.: National Institute of Justice.
- Sidebottom, A. & Tilley, N. (2011) *Further improving reporting in crime and justice: an addendum to Perry, Weisburd and Hewitt* (2010) *Journal of Experimental Criminology*.
- (2008) The Food and Drug Administration's standard for assessing the effectiveness of pharmaceutical drugs and medical devices, at 21 C.F.R. §314.126, retrieved from: <http://www.gpo.gov/fdsys/pkg/CFR-2008-title21-vol5/pdf/CFR-2008-title21-vol5-sec314-126.pdf>.
- Tierney, R. Mallender, J. (2015) (forthcoming) *The economic case for and against prison: update*. Optimity Matrix
- Urahn, S.K. (2012). The Cost-Benefit Imperative. Governing.com. Retrieved from: <http://www.governing.com/columns/mgmt-insights/col-cost-benefit-outcomes-states-results-first.html>.
- Villetaz, P. Gillieron, G. & Killias, M. (2015) "*The Effects on Re-offending of Custodial vs. Non-custodial Sanctions: An Updated Systematic Review of the State of Knowledge*", January 2015. Campbell Review retrieved from <http://www.campbellcollaboration.org/lib/project/22/>.
- White, H. (2013) *An Introduction to the use of randomized control trials to evaluate development interventions* International Initiative for Impact Evaluation, Working Paper 9.

Gloria Laycock is Professor of Crime Science at the UCL Department of Security and Crime Science. Her BSc and PhD were both in psychology. She was a researcher in the Home Office for many years, leaving as Head of the Home Office Police Research Group for a fellowship in the USA, before coming to UCL as Director of the Jill Dando Institute of Security and Crime Science. In 2011 she took special leave from UCL working in Abu Dhabi to establish a Community Policing and Police Science Institute for the Abu Dhabi Police.

Jacqueline Mallender is head of Public Policy for Optimity and a respected international economist specialising in the fields of justice, health and education. Jacqueline has a keen interest in the economics of crime and justice. She is a convenor of the Campbell Collaboration Crime and Justice Coordinating Group and the joint Campbell and Cochrane Economics Methods Group. Over the last 15 years, she has directed many evaluations of the effectiveness of interventions designed to reduce crime, reduce fear of crime and reduce offending. She has experience of crime prevention, policing, courts and sentencing, and correctional services. Her work has focused on the links between crime, alcohol and illicit drugs.