

Reviewing Evidence for Evidence-Based Policing

Kate Bowers, Lisa Tompson, Aiden Sidebottom, Karen Bullock and Shane Johnson

Evidence-based policing has widespread appeal. It calls for a shift from ways of working that are led by experience to that which is informed by the best available research evidence. Yet research evidence can, of course, take many forms and be of varying quality. In this chapter we consider the main sources of research evidence available to support decision making in the crime prevention field. We begin by discussing four main sources of research evidence: practitioner reports, primary evaluation studies, systematic reviews and reviews of reviews. We then focus on a sample of systematic reviews of single crime prevention interventions and, using the EMMIE framework (Johnson, Tilley and Bowers, 2015), we explore the extent to which these reviews adequately report information on the **E**ffect of intervention, the **M**echanisms through which interventions are believed to work, **M**oderators that may influence the impact of intervention, **I**mplementation issues that may impede or facilitate the intervention, and the **E**conomic costs of interventions. Based on this exercise, and similar efforts to review the evidence base for crime prevention, we next outline five features which we argue increase the value of evidence reviews in support of crime prevention. Finally, we discuss some practical steps to increase the likelihood of future evidence reviews incorporating these features.

Looking for evidence

In this section we review four sources of research evidence from the perspective of the practitioner, whose decision making the evidence base is ultimately intended to inform.

Practitioner reports

We start with what is likely to be most familiar to our practitioner, reports written by and for practitioners. Taking a variety of forms, we focus on those that involve the routine assessment and evaluation of crime prevention initiatives that practitioners may conduct themselves, or in collaboration with researchers. Such evaluations are conducted for a variety of reasons; for example, to meet the requirements of funders, to satisfy the host organisation/s that an investment was worthwhile, or as part of a problem-solving project. Examples include the Tilley Award entries in the UK (see Bullock *et al.*, 2006) and the Goldstein Award entries in the US (see Scott, 2000; Rojek, 2003). The Tilley and Goldstein Award schemes seek to acknowledge excellence in problem-oriented policing and require entrants to pay attention to evaluation. Consequently, they are especially useful in this context.

Practitioner reports are potentially good sources of information about the effectiveness of local level crime prevention interventions. They may contain information that is both useful and digestible to practitioners, but that might be overlooked by other evaluators. For example, Bullock *et al.* (2006) found material about implementation was available in just over 40 per cent of the 150 Tilley Award reports that they examined. This material provided useful lessons for practitioners hoping to avoid problems at the development and implementation stages of new projects.

There are two common problems with practitioner reports. The first relates to methodological quality. Research suggests that, at best, the methodological quality of practitioner reports is variable and oftentimes poor (Bazemore and Cole, 1994; Read and Tilley 2000; Scott, 2000; Bullock *et al.*, 2006). In discussing quality it is important to distinguish between two different forms of evaluation (see Eck, 2016). Process evaluations track the development and implementation of an intervention, and hence focus on what was done and how. Outcome

evaluations estimate the impact of an intervention on crime (or other outcomes), assuming the intervention has been implemented.

Process evaluations may lack necessary detail if practitioners have not understood the need to maintain detailed project records. They may see doing so to be too time-consuming, intrusive and peripheral to their day-to-day job (Bullock *et al.*, 2002; Forrest *et al.*, 2005). Whatever the reason, process evaluations often do not contain sufficient information to enable project replication. There are also recurrent problems with outcome evaluations. For example, it has been shown that practitioner reports often lack sufficient data, give little consideration to alternative explanations for observed effects, or to possible side-effects, and pay scant attention to how initiatives may have produced their effects (Read and Tilley, 2000; Bullock *et al.*, 2006).

A second problem is that although practitioner reports might be made available on an organisation's website, at conferences or through informal networks, they are unlikely to be visible or accessible to other practitioners. A notable exception are the above mentioned Tilley and Goldstein Award entries that are publically available by the US-based Center for Problem-oriented Policing (see www.popcenter.org), a web-based resource containing copious materials regarding what techniques are effective in reducing crime and in what circumstances, along with materials regarding the implementation of problem-oriented policing.

Academically motivated primary evaluation studies

Practitioners may draw on the results of academically motivated primary evaluations to guide their decision making; these being assessments of the impact of a crime prevention intervention at a given point in time, conducted by academic, government or other

professional researchers. Such evaluations might investigate the processes through which an intervention was delivered, demonstrable crime prevention effects, unintended consequences and cost effectiveness. To illustrate, two of us (Bowers and Johnson) conducted an impact evaluation of the effectiveness of fitting gates to networks of alleys running along the back of terraced properties to reduce opportunities for offending. We found that, relative to a reference area, burglary reduced by approximately 37% in the areas where alley gates were implemented, that there was a diffusion of crime control benefits to properties in the immediate surrounding areas, and that the scheme yielded savings of £1.86 for every £1 spent (Bowers *et al.*, 2004).

There are three central concerns about primary evaluation studies: quality, accessibility and interpretability. Firstly, addressing quality, evaluations vary in terms of their internal validity – the extent to which they are able to demonstrate cause and effect relationships (Campbell and Stanley, 1963). Many evaluations examine observed crime trends before and after an intervention is implemented but fail to consider whether any observed change was generated by the intervention or another (potentially unknown) factor. Threats to internal validity can be mitigated by establishing appropriate control groups. Superior to before and after evaluation designs, such quasi-experimental designs are, however, subject to selection bias. Areas or individuals assigned to treatment and control groups might differ in ways that might affect the outcomes observed.

Selection bias can be addressed by randomly allocating people or places to treatment and control conditions using a randomised control trial design (RCT), often seen as the ‘Gold Standard’ evaluation design (Sherman *et al.*, 1997). That said, where conducted properly, quasi-experimental studies can rule out most (if not all) threats to internal validity (Nagin and Weisburd, 2013; Eck and Madenson, 2009). Moreover, where interventions are tailored to the peculiarities of specific locations, as is called for in problem-oriented policing (see Goldstein,

1990), randomly allocating interventions to areas that did not experience the crime problems of interest would be of little value, even if the internal validity of the evaluation were high. Our aim here is not to debate which type of evaluation methodology is the best - this will vary according to the situation. Instead, our point is that methodological adequacy is important but varies across primary evaluation studies.

A second issue concerns the accessibility of primary evaluations. The results of academic studies are typically published in academic journals. Whilst access to academic journals is progressively changing due to open access publishing, many journals remain inaccessible to those not affiliated with a university (such as police practitioners) due to the expensive subscription fees. That the results of primary evaluations will be published at all, however, is not guaranteed. It is widely accepted that positive results are more attractive to the academic community, to funders and to journal editors (Franco *et al.*, 2014) and are therefore more likely to be published and appear in multiple outlets (Rothstein *et al.*, 2005). In contrast, studies yielding negative or null findings are less likely to be published and may never be submitted for publication – the so-called ‘file drawer problem’ (Rosenthal, 1979).

Fortunately, the issue of null findings is becoming more widely acknowledged in academia, and there are efforts to rectify this. For example, the *Journal of Experimental Criminology* now encourages the submission of high quality null finding evaluations in a short report format (personal communication from current Editor Lorraine Mazerolle)

A third concern relates to the interpretability of primary evaluations. All evidence, to some extent, requires interpretation: a judgement of its relevance to the presenting problem and context. It may be difficult for practitioners, who cannot realistically be expected to have the expertise necessary to interpret (often complex) statistical data, or esoteric academic jargon, to make sense of the findings of evaluations. Even where findings can be accessed and understood, interpreting the results of multiple studies, which may be contradictory, can be

difficult. Evaluation reports may or may not provide the sorts of contextual data that would inform practitioners about whether an intervention, demonstrably effective or otherwise, would likely be suitable for their local context and available resources. Considering the process of implementation, primary evaluations may or may not provide practitioners with the information necessary to successfully deliver an intervention (Bullock *et al.*, 2006).

Before moving on, it is worth rehearsing the different groups who might undertake primary evaluations. Some are generated by academics, others by practitioners, some by academic-practitioner partnerships and others still by consultancies or paid contractors. Each of these groups will have their own interests and incentives, and these should be considered by the reader. These interests will, for example, influence the degree to which the findings are produced by impartial advisors, the technical quality of the research, the likely medium of publication, and the financial resourcing of the exercise. For a more detailed account of some of these potential pitfalls see Tilley (1999). A key point is that insufficiently funded evaluations undertaken by researchers lacking in appropriate methods training can lead to both partial and misleading results.

Systematic Reviews

Keeping abreast of developments in the evaluation literature is difficult for busy practitioners. Systematic reviews are an important means of distilling the research evidence to support decision makers. They require that each step of the review process be conducted in a transparent way so as to enable future replications. The process of conducting a review is characterised by explicit objectives and a rigorous searching and screening process whereby primary studies are accepted or rejected on the basis of an explicit eligibility criteria, with exclusions justified and documented (Farrington *et al.*, 2001). The quality of the candidate

studies are usually critically appraised before the data from them are marshalled to produce new knowledge that is 'greater than the sum of the individual studies' (Gough and Thomas, 2012: 39).

Where the data from primary studies are sufficiently similar, and it is statistically appropriate, quantitative methods such as meta-analysis can be used to produce estimates of the overall effectiveness of an intervention. A meta-analysis weights the effect sizes of individual studies by their respective sample sizes and variance (i.e. the reliability of the estimated effects), thus producing a pooled effect size (and estimate of the precision of that statistic) that reflects the reliability of the findings. In combining samples across multiple studies, meta-analysis increases statistical power, thus enabling an overall effect of an intervention to be detected (should one exist). This is a key advantage of this method, and one reason why systematic reviews with meta-analyses are the dominant source of evidence in other allies of evidence-based practice, most notably medicine. Pooling data from multiple primary evaluation studies provides a quantitative summary of the evidence as it relates to a given topic at a set point in time and, crucially, absolves practitioners of the tricky task of making sense of sometimes contradictory evidence from different studies.

An informative meta-analysis will also explore those factors which might feasibly *moderate* the effect of an intervention. That is, the meta-analyst will specifically look to code information on factors deemed to influence the outcomes produced by the intervention under review. For example, does CCTV work better in car parks or residential areas, or does cognitive-behavioural therapy work better for male or female offenders? This information can then be used to partition the studies into meaningful sub-groups and pooled effect sizes can be computed for each, revealing whether these factors are relevant to the observed effect. More sophisticated analytical techniques (e.g. meta-regression, multilevel modelling or

Bayesian analysis) can be used to make causal inferences about the degree to which these factors moderate the effect.

Systematic reviews are important for communicating high-level research findings to practitioners and policy-makers. Yet, there are good reasons why in crime reduction, unlike in medicine, systematic reviews are not a mainstay of professional practice. First, high-quality primary evaluations are rare and for some topics virtually absent (e.g. organised crime, terrorism, modern slavery). While ‘empty’¹ systematic reviews may be useful for researchers, they are of little utility to practitioners. Second, there is a tendency for systematic reviews to focus on studies published in scientific journals. Failure to consult the ‘grey literature’ of government and practitioner reports, which comprises a sizable proportion of the crime reduction literature (Wilson, 2009; Tompson and Belur, 2015), may give rise to biased findings and neglect potentially important information. Finally, the methods used to quantitatively synthesise data from primary studies can be complex, and effectively communicating these methods to a lay-audience is challenging. The inclusion of ‘plain English’ executive summaries is now increasingly common in a bid to help overcome this issue.

Reviews of reviews

The next section of this chapter will discuss strategies for reviewing the quality of evidence, focussing in particular on systematic reviews. For completeness, however, it is important to acknowledge that ‘reviews of systematic reviews’ or ‘meta-reviews’ are an additional source

¹ An ‘empty’ systematic review is one in which no publications meet the criteria outlined by the authors for inclusion. Typically, the result of this includes the inability to perform a meta-analysis which statistically summarises the average effectiveness of an intervention.

of research evidence (Gough and Thomas, 2012) that might plausibly inform crime prevention decision making. Reviews of reviews can take several forms. They may act as an umbrella review on a particular topic (for example the prevention of drug use in young people, see Canning *et al.*, 2006), a particular approach to crime reduction (see for example Bowers and Johnson 2016 who undertake a review of reviews employing situational crime prevention measures) or as a way of summarising existing reviews across an entire field (such as the ‘Preventing Crime: What Works, What Doesn’t, What’s Promising’ by Sherman *et al* 1997).

Reviews of reviews have their own strengths and weaknesses. They have the significant advantage of being a ‘one-stop-shop’ for busy practitioners and academics. They can also usefully present evidence in a consistent format across a number of topics, making the comparison of ideas and outcomes much quicker and easier. Ultimately, their value relies on the quality of the research reports that they draw on and in this case they are further removed from the original context- they depend on the systematic reviewer’s interpretation of the original primary evaluations, which of course might be biased or incomplete. This is why it is essential to consider the quality of the underlying evidence, to which subject we now turn.

To summarise the types of evidence discussed in this section, and to document the key features of each in terms of strengths and weaknesses, we have compiled Table 1 for reference.

Table 1. The strengths and weaknesses of different research evidence for practitioners

Source of research evidence	Main purpose	Main strengths	Main weaknesses
Practitioner reports	To convey local knowledge of crime prevention initiatives	<ul style="list-style-type: none"> • Written in accessible language for practitioners • Can contain contextual information relevant for implementation 	<ul style="list-style-type: none"> • Often use weaker outcome evaluation designs that lack internal validity (e.g. pre- post-intervention designs) • Process evaluations can be incomplete as detailed information not collected/recorded • Little consideration may be given to alternative explanations for observed effects, or to possible side-effects • Scant attention paid to how initiatives may have produced their effects (mechanisms)
Academically motivated primary evaluation studies	To report the impact of initiatives to academic audiences	<ul style="list-style-type: none"> • <i>May</i> be methodologically robust (i.e. have good internal validity) • Will have been critiqued by other academics 	<ul style="list-style-type: none"> • Publishing outlets can be inaccessible • Negative, or null, results may never be published • Language may be inaccessible
Systematic evidence reviews	To synthesise evidence on a given initiative	<ul style="list-style-type: none"> • Arguably the most reliable form of evidence • Consider the composite findings of an evidence base • Pool the effect to generate an overall conclusion 	<ul style="list-style-type: none"> • Are dependent on the quality of primary evaluations • May not include ‘grey literature’ (i.e. non-academic sources) • Language may be inaccessible
Review of reviews	To synthesise evidence across broad topics	<ul style="list-style-type: none"> • Act as a ‘one-stop-shop’ for practitioners • Consistently display evidence from multiple sources • Can identify quality and evidence gaps 	<ul style="list-style-type: none"> • Are dependent on the quality of systematic reviews which are in turn dependent on the quality of primary evaluations. • Given their breadth such exercises may need frequent updating to remain relevant.

Grading evidence

We now turn our attention to efforts to grade research evidence, and take stock of evidence rating tools for crime reduction. Apparent in many definitions of evidence-based practice is the notion that some evidence is better than others, and that decisions should be informed by the ‘best’ available evidence. For Sackett *et al.* (1996: 3) ‘*Evidence-based medicine ... is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients*’. For Sherman (1998: 3) ‘*Evidence-based policing is the use of the best available research on the outcomes of police work to implement guidelines and evaluate agencies, units, and officers*’.

Distinguishing ‘better’ from ‘worse’ evidence is a key requirement for evidence-based practice. To meet this aim, tools have been established to rate the quality of research evidence, and the confidence that can be placed in recommendations derived from it. The Canadian Task Force on the Periodic Health Examination is widely credited with producing the original ‘hierarchy of evidence’ in 1979. Like many subsequent evidence rating scales, it was oriented towards questions of intervention effectiveness (did it work?) and the risk of bias associated with different sources of evidence. RCTs thus received the highest grade, as they are widely considered to be the most trustworthy source of primary evaluation evidence.

Numerous evidence rating systems have since been developed to meet the requirements of specialist fields. A 2002 systematic review (West *et al.*, 2002) identified over 100 evidence rating systems for health care research, pertaining (in descending order) to RCTs (n = 49), systematic reviews (n = 20), observation studies (n = 19) and diagnostic test studies (n = 18).

Until recently, the only evidence rating scale commonly applied to criminological interventions was the Maryland Scientific Methods Scale (SMS). The SMS emerged out of a review of the evidence base for crime prevention (Sherman *et al.*, 1988). It is a simple five

point scale designed to assess internal validity and communicate the results to consumers of that evidence. A summary of the SMS is given below as Table 2:

Table 2. Maryland Scientific Methods Scale for rating crime prevention evaluations.

Least rigorous	1. Correlation between a crime prevention program and a measure of crime or crime risk factors.
	2. Temporal sequence between the program and crime or risk outcome clearly observed, or a comparison group present without demonstrated comparability to the treatment group.
	3. A comparison between two or more units of analysis, one with and one without the program.
Most rigorous	4. Comparison between multiple units with and without the program, controlling for other factors, or a non-equivalent comparison group has only minor differences evident.
	5. Random assignment and analysis of comparable units to program and comparison groups

According to the SMS, studies that report only a correlation between crime levels and an intervention are deemed ‘weak’ in terms of causal validity. For example, studies that show that the risk of crime is lower in areas that have Neighbourhood Watch (relative to those that do not) would reach level one. Where there is a ‘temporal ordering of cause and effect’, evidence is deemed ‘moderate’. In our Neighbourhood Watch example, a study would meet this criterion if levels of crime were tracked for the period pre- and post-intervention. The use of control groups provides greater evidence of causality, but the type of control group matters. The random allocation of units (e.g. areas or people) to treatment and control groups is designed to eliminate selection bias and, if done correctly, leads to equivalence across groups, meaning that rival (null) hypotheses can generally be ruled out. According to the SMS, RCTs are the most rigorous type of evaluation design. Alternative methods of selecting control groups, such as matching areas of intervention to areas with comparable social or demographic characteristics, would receive a moderate rating.

The SMS was proposed as an intuitive, easily communicable method of appraising methodological quality in primary studies. The architects of the SMS do, however, acknowledge a number of limitations. Farrington, Gottfredson, Sherman and Welsh (2002) discuss how the SMS is silent on 1) the mechanisms through which an intervention might bring about the observed effects, 2) external validity (the extent to which the observed findings are generalisable) and 3) that certain research designs – which may have strong internal validity (e.g. multiple time-series designs) – are not included.

The US National Institute of Justice has launched a more recent initiative to systematically grade the research evidence to support policing and crime prevention at CrimeSolutions.gov. This aims to identify and rate the available evidence on what works to prevent crime and offending. CrimeSolutions.gov focuses on interventions (such as CCTV) rather than problems (such as burglary). For interventions that do not yet have systematic reviews, the methodological adequacy of existing primary evaluations is assessed for rigor and the extent to which interventions were implemented as intended. Findings are then collated to provide an indication of whether the intervention has been shown to work, and the strength of the evidence on which this assessment is based. Where systematic reviews do exist, the quality of the review evidence is assessed. In both cases, studies are assessed by expert reviewers using standardised scales that draw on (but are not limited to) the SMS. By presenting findings using a simple evidence rating scale in combination with easy to read narratives that describe the intervention, evaluation outcomes, issues associated with implementation (and where available the costs), the aim is to provide practitioners and policy makers with an understanding of what the evidence base suggests.

A similar initiative is underway in the UK as part of the What Works Centre for Crime Reduction, hosted by the College of Policing. Work conducted to date has sought to systematically identify, rate and rank existing systematic reviews of crime prevention

interventions, as well as carry out new systematic reviews to fill knowledge gaps where existing reviews are out of date, limited in scope or otherwise do not sufficiently inform crime prevention practice. In the next section we describe the evidence rating scale – EMMIE – generated by What Works Centre for Crime Reduction, and report the results of applying this scale to a sample of systematic reviews.

Applying EMMIE to grade systematic review evidence

CrimeSolutions.gov and the UK What Works Centre have broadly similar objectives: that of identifying, grading and synthesizing research evidence in a manner that best supports the needs of decision makers. An important difference between the approach of the UK What Works Centre and other similar exercises is that systematic reviews are systematically rated not just on methodological adequacy – which is, of course, vital – but along other dimensions that are also important to practitioners and policy makers. As mentioned in the introduction, the EMMIE framework was designed to gather evidence (or acknowledge the absence of it) on five key dimensions (encapsulated by the acronym ‘EMMIE’) – (a) the **E**ffect of the intervention, (b) the causal **M**echanism(s) through which interventions are intended to work, (c) the factors that **M**oderate intervention effectiveness, (d) the articulation of practical **I**mplementation issues, and (e) the **E**conomic costs of intervention (Johnson *et al.*, 2015).

These five elements of EMMIE were selected to ensure that reviews do not just focus on quantitative outcomes (‘Effect’) following the intervention of a crime prevention scheme. This limited focus fails to acknowledge that the impact of social interventions may vary substantially in different contexts (these act as ‘moderators’ for likely outcomes) and when different mechanisms are ‘fired’ (for example the use of CCTV as a visible deterrent in a car park compared to the submission of CCTV evidence to convict for a common assault in a town centre). Implementation is distinct from, but related to, these issues and continues to be

highly challenging in this field (see e.g. Knutsson and Clarke 2006). This is because gathering and mobilizing agencies and individuals; gaining appropriate funding; accessing the correct tools and equipment; garnering public support, consent and involvement; dealing with administrative red tape; and negotiating unanticipated issues, are all complex and time consuming undertakings. Finally, information on cost effectiveness and cost benefit are of huge interest to practitioners and policy makers, but collecting detailed information on both the inputs to (for example staff time, equipment costs, consumables) and outputs of a crime prevention scheme (number of hours delivered, number of physical measures installed) is rarely done in the sort of systematic way that would be of use to others planning similar exercises.

A coding instrument was devised to systematically rate systematic reviews along the dimensions of EMMIE (see Tompson *et al.*, 2015). This instrument included around 100 fields on which evidence was either extracted (the EMMIE-E codes) or quality-graded (the EMMIE-Q codes). As part of the What Works Centre for Crime Reduction, this instrument has been used to appraise the quality of 87 systematic reviews on crime prevention (for details on how the reviews were identified see Bowers *et al.*, 2014). Here we present the results for a subset of police-relevant reviews (n=6) to illustrate the quality of evidence pertaining to interventions and strategies that the police might reasonably be able to implement, whilst acknowledging that a far wider spectrum of crime prevention interventions exist that can be implemented by, or with, other agencies. The topics of the systematic reviews were (in descending order): sobriety checkpoints (3); hot spots policing (1); second responder programmes (1); and street-level drug law enforcement (1).

The quality scores for Effect varied across the six systematic reviews, although four reviews were seen to score three or four, indicating that they had sufficiently considered elements of validity that might bias the reliability of the effect reported (Table 3). Mechanism was less well covered. Four of the six reviews made a broad statement of how the intervention was

believed to work, one did not mention mechanisms at all, and one provided a detailed review. Moderators were covered a little more thoroughly. Four of the SRs used existing theory (or empirical evidence) to describe or test factors that could moderate the effect of the initiative. Efforts to document implementation factors that might impede or facilitate a crime prevention initiative were rare, with only one SR doing so in an evidence-based manner (actually providing quantitative or qualitative information on what happened), and most systematic reviews provided only ad-hoc comments on this dimension (and hence inconsistently documented evidence across evaluations). Lastly, information on economics was entirely absent from these reviews. This is by no means atypical for crime prevention studies, an exercise on a larger number of systematic reviews in the field demonstrated that there is a significant paucity of financial information reported (Johnson et al 2016).

Table 3. The quality scores for six policing and partnership SRs on each EMMIE dimension

EMMIE dimension	Quality score	N reviews
Effect	1. Considered no elements of validity	0
	2. Considered 1 element of validity	1
	3. Considered 2 elements of validity	1
	4. Considered 3 elements of validity	2
	5. Considered 4 or 5 elements of validity	2
Mechanism	0. No mention	1
	1. Broad statement	4
	2. Detailed review	1
	3. Formal model and predictions	0
Moderator	4. Tested using data	0
	0. No mention	0
	1. Ad-hoc description	1
	2. Post-hoc test of moderators	1
Implementation	3. Theory-based description	2
	4. Theory-led data analysis	2
	0. No account	0
	1. Ad-hoc comments	4
Economics	2. Systematic efforts to document	1
	3. Detailed evidence-based account	1
	4. Complete evidence-based account	0
	0. No mention of costs	6
	1. Only direct costs/benefits estimated	0
	2. Direct and indirect costs/benefits estimated	0
	3. Marginal/opportunity costs	0
	4. Costs by bearer	0

Requirements of 'good' systematic evidence reviews

In this section we discuss what can be gleaned from the coding exercise presented above and translate these lessons into recommendations for what a 'good' systematic evidence review should include, taken here to mean useful to decision-makers.

1. Evidence reviews need to cover 'EMMIE'

The previous section highlights the need for systematic reviews in crime prevention to collect and report evidence (or the lack of it) on each of the dimensions of EMMIE. We therefore recommend that each of these elements is considered by those producing systematic reviews in the field. Whilst this will make the coding of reviews more intensive, we believe the gain to crime prevention practice would be significant. Hence, in principle, instead of purely knowing whether a given intervention has been shown to be effective in preventing crime, practitioners will have knowledge about how and where an intervention is most likely to work, what the implementation challenges are likely to be and information on expected costs. We contend that this additional information will enable decision makers to make better judgements about the potential for success given their particular local conditions. Note we present this both as an aspiration and as a method for organising existing information. A big hurdle to achieving this goal is the observation that primary studies seldom report information beyond effect in great detail (see Sidebottom *et al.*, 2015) and that standard systematic reviews often select studies largely based on research design which may preclude studies with useful non-effect information from being included.

Fortunately, the lack of information on possible mechanisms and moderators in systematic reviews is beginning to be more widely acknowledged and addressed. For example, the

Campbell Collaboration now seeks reviews to be more explicit in terms of mechanisms and logic models (see Policies and Guidelines²). A recent example is a systematic review of candidate mechanisms to explain the effect of broken windows policing undertaken by Weisburd et al (2015). Combined with renewed pressure from academics such as Nancy Cartwright and colleagues (Cartwright and Hardie, 2012) for evaluation studies to collect information beyond just intervention effectiveness, we are cautiously optimistic that reviews with a wider breadth of evidence will begin to become available.

2. Evidence reviews should make the ‘holes’ explicit and assess the quality of what is available

EMMIE offers the opportunity to systematically rate a wider range of evidence types. For example, when reporting information about the calculation of an effect size, detailed guidance on how to report the results of systematic literature searches, explaining the statistical meta-analysis procedure taken, assessing the reliability of results given the extent of the evidence used and reporting any biases in the procedure (such as coder bias or publication bias) is already available (see Gwet, 2014; Duval and Tweedie, 2000; Egger, 1997). However, once assimilated across primary studies, assessing the quality of implementation information, descriptions of the various contexts and likely mechanisms is not commonly done. The EMMIE coding tool encourages quality appraisal across a fuller range of relevant dimensions, rather than assigning primacy to just the ‘effect’ of an intervention.

² The Campbell Collaboration both commissions and provides a library of systematic reviews of the research evidence in social policy including criminal justice.
http://www.campbellcollaboration.org/artman2/uploads/1/C2_Reviews_policy__guidelines_draft_5-1-13.pdf

One question is whether the assessment of information along these lines should be an activity undertaken by systematic reviewers themselves, given their knowledge of the primary evidence available, or by those at a policy level who are assessing the quality of competing systematic reviews. We would argue that assessment at both these levels is useful – the former would make the latter more expedient and would also ensure that those undertaking reviews are cued to look for various pieces of information in primary studies and to do this with a critical eye. Neither should the quality of the systematic review and the quality of the primary evidence be conflated. Thorough reviews can come back with ‘nil returns’ and this is useful to acknowledge in quality assessments.

As discussed above (in Table 3) and elsewhere (Bowers *et al.*, 2014) research conducted for the UK College of Policing has demonstrated that there are considerable holes in the evidence base on crime prevention. Holes result from no evidence being available in primary studies and from systematic reviews that do not report information where it is available. It would be useful if systematic reviews made this distinction clear. Use of EMMIE would encourage such reporting.

3. Reviews should summarise available primary studies for practitioners

There is wide variation in the degree to which reviews give contextual details or narratives about each of the evaluations that they summarise. Some include a basic ‘one-liner’ for each intervention, others include tables summarising the (reviewer-deemed salient) features of primary studies whilst others provide longer narrative-based summaries. The latter give textual detail but take up space and might not record primary evaluation evidence systematically and accessibly. We recommend a systematic way of reporting narrative or contextual information on primary interventions, presented in a format where easy comparisons are possible. A possibility is to apply EMMIE at the *primary* study level to do

this. Each primary study could be summarised on the 5 key elements and a summary table to enable comparison across primary studies. Structured narratives could additionally be added. These have been made available at the systematic review level for the What Works Centre Crime Reduction Toolkit and could be sensibly applied to primary evaluations as well (see <http://whatworks.college.police.uk/toolkit/Pages/Toolkit.aspx>).

4. Reviews should be written with practitioners in mind

We have mentioned that some academic products can be difficult for practitioners to interpret. Good examples of accessible academic writing for practitioners include the Pop Center problem-solving and response guides. These synthesise evidence from multiple sources (practitioner and academic reports) in a way that speaks directly to practitioner needs. These are not conducted systematically, as defined herein, but are thorough and comprehensive. Working outside these conventions means they do not require the academic jargon of formal systematic reviews – which makes them immediately more accessible, allows more room for the inclusion of tacit knowledge (Tilley, 2006) or ‘golden nuggets’ of information from many sources (independent of the *overall* quality of the original evidence) and affords more flexibility in their presentation. Some guides focus on interventions (or responses), others on particular crime or disorder problems. COPs problem-solving guides include summaries of each response (as part of an appendix outlining the different possible approaches) setting out whether it appears to work, how it works, the conditions under which it works best and any considerations that are necessary in terms of potential barriers to effective implementation.

This speaks to our earlier argument that more than the ‘effect’ needs to be considered and highlights the final point we will make here: that *summaries* really matter. Many practitioners will only read a summary so it really needs to contain the key information and be clear. It is

for this reason that there are often guidelines for systematic review summaries – such as those offered by Campbell Collaboration – although arguably these have a tendency to focus on methods and quantitative effect sizes rather than contextual information for practitioners. Fortunately academics are becoming more mindful of difficulties with communicating key messages concerning evidence to practitioners. The principles associated with ‘translational criminology’ are becoming more generally understood and recognised. Translational criminology is a field in its own right which seeks to translate scientific discoveries into policy and practice through studying knowledge dissemination (Laub 2011).

5. Reviews should be amenable to being integrated into knowledge ‘hubs’

When focusing on the search for the best intervention with which to address a specific crime problem, practitioners might need to review the potential of multiple interventions from several systematic reviews. When constructing a systematic review, thought could usefully be given to how easily the evidence from one systematic review can be integrated with reviews of potential alternatives.

A key activity of What Works Centre has been to summarise information at the intervention level in a consistent way (based on EMMIE) across different systematic reviews. As mentioned above, although other exercises in integrating this type of information exist, we focus on our experiences with the production of the on-line Crime Reduction Toolkit (whatworks.college.police.uk/toolkit/). Some reviews were more easily streamlined in the integration process than others; for example, comprehensive reviews and those that used the ‘common language’ of crime prevention practice (see Ekblom, 2010). Conversely, in some reviews, even basic information about where and when the intervention took place were hard to find or missing. Similarly, across countries and places descriptions of the actors involved can vary (for example ‘PCSOs’ in the UK are a type of police officer with specific duties and

powers). Reviews written in languages other than English are often missed in such integration exercises. Finally, for reviews that do focus on problems rather than interventions, and therefore review multiple approaches, it is often difficult to distinguish the specific information at the intervention-level that would be necessary for integration into a knowledge hub organised by intervention.

Some suggestions on practical steps to improve the value of evidence reviews for crime prevention

We have listed five features which we believe would increase the practical value of evidence reviews for crime prevention. Our suggestions to encourage future reviews to incorporate these features are as follows:

1. Develop reporting guidelines for crime prevention reviews.

Presently, there is no agreed-upon standard for the reporting of systematic reviews in criminology. According to the Enhancing the QUALity and Transparency Of health Research (EQUATOR) network there are now over 250 reporting guidelines in the medical sciences. A direct response to the seemingly pervasive inadequacies of much research reporting which threaten to stifle replication and waste resources, these guidelines aim to improve the completeness and consistency of reporting.

It is not that criminologists are ignorant to the existence of reporting guidelines or that they are immune to the sorts of shortcomings they are designed to overcome. Far from it. In their appraisal of 62 published RCTs using the CONSORT statement, Perry, Weisburd and Hewitt (2011) identified several recurrent omissions and reporting deficits, with over half of the reviewed studies receiving a 'low' rating for the quality of reporting.

Guidance on what should be included when reporting a systematic review is already available in the form of the PRISMA statement. PRISMA originated in medicine, however, and therefore does not fully capture all the issues that are important in the reporting (and conduct) of systematic reviews in criminology (see Sidebottom and Tilley, 2012). EMMIE provides a convenient framework around which to generate reporting guidelines for crime prevention reviews, in addition to modifying those reporting guidelines already available.

2. Encourage Supplementary online materials.

The type of evidence review which we are advocating likely contains more words than a standard evidence review. This might not sit well with academic journals, the main outlets for systematic reviews. To address this issue we might take the lead from other disciplines which increasingly allow for supplementary material to be attached to a published article.

3. Lobby researchers, practitioners and policymakers to speak EMMIE

Ultimately we need to make sure practitioners engage with the drive to improve the evidence base. Much work is still required to cultivate a culture in which research evidence that speaks to the dimensions of EMMIE is a) routinely collected and reported by evidence producers (i.e. researchers) and b) routinely and explicitly sought by evidence consumers (i.e. practitioners and policymakers). There are a number of fronts on which activity is required here. Practitioners need to (a) know where to look for evidence (b) have an appetite to use it (c) understand it sufficiently to use it successfully in evidence-based practice (d) be aware of the current limitations of what is available and (e) be motivated to collaborate in exercises to improve evidence. This is a tall order, but is one that we believe we should be collectively moving towards.

Conclusion

This chapter has considered the role of research evidence in support of crime prevention decision-making. We have highlighted different forms of research evidence and discussed their respective strengths and weaknesses. We explained the role that research design plays in determining the reliability of evaluation evidence. In considering the task of grading the quality of research evidence both in crime prevention and in other fields, we argued that the majority of such exercises concentrate only on rating the quality of the process of generating the effect size (outcome). Whilst essential, we argue that this does not go far enough in speaking to other key requirements for good practice. Information on how interventions work, for whom, under what conditions, with what input and at what cost needs to not only be systematically reported but also rated for quality. We need to mobilise the crime prevention community to see the value of doing this and to engage in exercises to fill the existing gaps in the evidence base.

References

- Bazemore, G. and Cole, A. (1994). 'Police in the Laboratory of the Neighbourhood: Evaluating Problem-Oriented Strategies in a Medium Sized City', *American Journal of the Police* 13(3): 119-48.
- Black, N. and Donald, A. (2001). 'Evidence based policy: proceed with care commentary: research must be taken seriously', *BMJ*, 2001, 323: 275-79.
- Bowers, K., Johnson, S. and Hirschfield, A. (2004). 'Closing Off Opportunities for Crime: An Evaluation of Alley-Gating,' *European Journal on Criminal Policy and Research*, 10(4): 285-308
- Bowers, K.J. Tompson, L and Johnson, S.D. (2014). 'Implementing information science in

- policing: mapping the evidence base', *Policing: A Journal of Policy and Practice*, 8(4): 339 - 52. doi: 10.1093/police/pau052.
- Bullock, K., Farrell, G. and Tilley, N. (2002). Funding and Implementing Crime Reduction Projects, *Home Office Online Report 13*, London: Home Office.
- Bullock, K., Erol, R., and Tilley, N. (2006). *Problem-oriented policing and Partnership*, Cullompton: Willan Publishing.
- Campbell, D.T. and Stanley, J.C. (1963). 'Experimental and quasi-experimental designs for research on teaching, in N.L. Gage (ed.), *Handbook of research on teaching*, pp. 171–246. Chicago, IL: Rand McNally.
- Cartwright, Nancy and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing it Better*. New York: Oxford University Press.
- Duval, S. and Tweedie, R. (2000). 'Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis', *Biometrics*, 56(2): 455-63.
- Eck, J.E. (2016). Evaluation for lesson learning. In N. Tilley and A. Sidebottom (eds), *Handbook of Crime Prevention and Community Safety* (2nd eds). Routledge.
- Eck, J.E. and Madensen, T. (2009). 'Using Signatures of Opportunity Structures to Examine Mechanisms in Crime Prevention Evaluations,' in J. Knutsson and N. Tilley (eds) *Evaluating Crime Reduction Initiatives*, *Crime Prevention Studies*, vol. 24, pp. 59-48. Monsey, NY: Criminal Justice Press.
- Egger, M., Davey Smith, G., Schneider, M. and Minder, C. (1997). 'Bias in meta-analysis detected by a simple, graphical test', *British Medical Journal* 315(7109): 629–34.
- Eckblom, P. (2010). *Crime prevention, security and community safety using the 5Is framework*, UK: Palgrave Macmillan
- Farrington, D.P., Petrosino, A., and Welsh, B.C. (2001). 'Systematic Reviews and Cost-Benefit Analyses of Correctional Interventions,' *The Prison Journal*, 81(3): 339–59.

Farrington, D., Gottfredson, D., Sherman, L. and Welsh, B. (2002). 'The Maryland Scientific Methods Scale,' in D. Farrington, D. MacKenzie, L. Sherman and B. Welsh (eds) *Evidence based Crime Prevention*, London. Routledge, pp 13-21.

Forrest, S., Myhill, A. and Tilley, N. (2005). Practical lessons for involving the community in crime and disorder problem-solving, *Home Office Development and Practice Report, 43*, London: Home Office.

Franco, A. Malhotra, N. and Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer, *Science* 345(6203): 1502-05.

Gough, D. and Thomas, J. (2012). 'Commonality and Diversity in Reviews', in D. Gough, S. Oliver, and J. Thomas (eds) *An introduction to systematic reviews*, (pp. 35–65), London: Sage Publications Ltd.

Gwet, K. L. (2014). *Handbook of Inter-Rater Reliability*, Fourth Edition, Gaithersburg: Advanced Analytics, LLC.

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

Johnson, S.D., Bowers, K.J., Tompson, L., Laycock, G., Tilley, N., and Sidebottom, A. (2016). What Works to Reduce Crime, How does it work, and what else do we need to know? Academy of Criminal Justice Sciences, Denver, USA, April 2016

Johnson, S.D., Tilley, N., and 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

Knutsson, J. and Clarke, R.V.G. (eds) (2006). Putting Theory to Work, *Crime Prevention Studies*, vol. 20,163-197, Monsey NY: Criminal Justice

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis J.P., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration, *Journal of Clinical Epidemiology*, 62:1–34.

- Laub, J. H. (2011) Strengthening NIJ: Mission, Science and Process. *NIJ Journal / Issue No.* 268 October 2011
- Nagin, D. and Weisburd, D. (2013). 'Evidence and Public Policy: The Example of Evaluation Research in Policing', *Criminology and Public Policy*, 12, 651-79.
- Perry, A.E., Weisburd, D., and 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(3), 245-62.
- Raynor, P. (2008). 'Community penalties and Home Office research: On the way back to 'nothing works'?', *Criminology and Criminal Justice*, 8(1): 73-87.
- Read, T. and Tilley, N. (2000). Not Rocket Science? Problem Solving and Crime Reduction, *Crime Reduction Research Series Paper 6*, London: Home Office.
- Rojek, J. (2003). 'A Decade of Excellence in Problem-Oriented Policing: Characteristics of the Goldstein Award Winners', *Police Quarterly*, 6(4): 492-515.
- Rosenthal, R. (1979). 'The file drawer problem and tolerance for null results', *Psychological Bulletin*, 86(3): 638-41.
- Rothstein, H.R., Sutton, A.J. and Borenstein, M. (eds) (2005). *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, Chichester, UK: John Wiley & Sons, Ltd.
- Sackett D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B. and Richardson, W.S. (1996). 'Evidence based medicine: what it is and what it isn't', *BMJ* 1996; 312:71
- Sidebottom, A. and Tilley, N. (2012). 'Further Improving Reporting in Crime and Justice: an addendum to Perry, Weisburd and Hewitt (2010)', *Journal of Experimental Criminology*, 8(1): 49-69.
- Sidebottom, A., Tompson, L., Thornton, A., Bullock, K. Tilley, N., Bowers, K. and Johnson, S.D. (2015). *Gating alleys to reduce crime: a meta-analysis and realist synthesis*. Report submitted to the College of Policing. Available online at http://whatworks.college.police.uk/About/Documents/Alley_gating.pdf

Canning, U., Millward, L., Raj, T., and Warm, D. (2004). Drug use prevention among young people: a review of reviews. London: Health Development Agency.

Scott, M. (2000). *Problem-Oriented Policing: Reflections on the First 20 Years*, Washington D.C.: US Department of Justice, Office of Community Oriented Policing Services.

Sherman, L.W. (1998). 'Evidence-based policing', *Ideas in American Policing*, Washington, DC: Police Foundation.

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

Tilley, N. J. (2006). 'Guidance and good practice in crime prevention, in R. Clarke and J. Knutsson (eds) *Putting Theory to Work*, Criminal Justice Press: Monsey, NY.

Tompson, L., Bowers, K.J., Johnson, S.D., and Belur, J.B. (2015). EMMIE evidence appraisal coding tool, UCL Department of Security and Crime Science: London, UK, <http://discovery.ucl.ac.uk/1462093/>

Tompson, L., and Belur, J. B. (2015). 'Information retrieval in systematic reviews: a case study of the crime prevention literature', *Journal of Experimental Criminology*. doi:10.1007/s11292-015-9243-x.

Tilley, N. J.(1999) Project evaluation: problems and pitfalls. In: Marlow, A and Pearson, G, (eds.) *Young People, Drugs and Community Safety*. Russell House: Lyme Regis.

Timmermans, S. and Berg, M. (2003). *The gold standard: the challenge of evidence-based medicine and standardization in health care*, Temple University Press.

Weisburd, D. (2000). Randomized Experiments in Criminal Justice Policy: Prospects and Problems, *Crime & Delinquency*, 46(2): 181-93.

West S., King V., Carey T.S., et al. (2002). *Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47* (Prepared by the Research Triangle Institute–University of North Carolina Evidence-based Practice Center under

Contract No. 290-97-0011). AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality. April 2002.

Wilson, D.B. (2009). 'Missing a critical piece of the pie: simple document search strategies inadequate for systematic reviews', *Journal of Experimental Criminology*, 5(4): 429–40.