

On the development and application of EMMIE: Insights from the What Works Centre for Crime Reduction

Abstract

The *What Works Centre for Crime Reduction* was established in September 2013 with the aim of increasing the use of research evidence by decision-makers in policing and crime reduction. The EMMIE framework was developed to meet this aim. It encapsulates five broad categories of research evidence that are considered relevant to crime reduction, namely Effect size, Mechanism, Moderator (or context), Implementation and Economics. In this paper, we chart the origins and development of EMMIE. We also reflect on our experience of applying EMMIE both as a coding system to appraise systematic review evidence and as a framework to inform the design and conduct of systematic reviews in crime reduction. We conclude with a critique of EMMIE and with suggestions on how it might be developed and refined in the future.

Key words: crime reduction, EMMIE, evidence-based policing, systematic review

Background

In the past two decades, there has been increased interest in evidence-based approaches to policing and crime reduction (see Sherman, 1998, 2013; Knutsson and Tompson, 2017; Lumsden and Goode, 2016). It is a movement that few have challenged¹. Calls for decisions on how to deal with crime to be informed by reliable research evidence is both widely advocated and accepted. The UK Government has invested heavily in promoting and facilitating evidence-based practice. In 2013, it launched the ‘What Works Network’, a national initiative aspiring to “improve the way government and other organisations create, share and use high quality evidence for decision-making”². To this aim, at the time of writing there are seven What Works Centres and two affiliate members covering a range of policy areas including ageing, education, health and, most relevant to this article, crime reduction.

The focus of this paper is the *What Works Centre for Crime Reduction* (WWCCR), and in particular the EMMIE framework created by Johnson, Tilley and Bowers (2015) to support the development of the WWCCR. We begin by describing the origins of and rationale for the EMMIE framework. The proposed functions of EMMIE are then outlined, as they relate to both primary research and evidence syntheses in crime reduction. Next, we discuss previous attempts to use EMMIE as a framework with which to assess the extent and quality of systematic review evidence in crime reduction, which reveal a general failure to report evidence on the implementation of crime reduction interventions and their expected/realised economic costs and benefits. We then reflect on efforts to use EMMIE as a guide to inform the

¹ This is not to ignore critiques of evidence-based policing as it is generally conceived (see for e.g. Eck, 2017b, Sparrow, 2016; Tilley and Laycock, 2017).

² <https://www.gov.uk/guidance/what-works-network>

conduct and reporting of systematic reviews in crime reduction. Finally, we consider the next steps in the usage and evolution of EMMIE.

On EMMIE: Origins and Development

The *What Works Centre for Crime Reduction* was established in September 2013. It has been supported by a consortium of British universities led by University College London alongside Cardiff University, Surrey University, Dundee University, Birkbeck, the Institute of Education, University of Southampton and the London School of Hygiene and Tropical Medicine (see Gough, Maidment and Sharples, 2018, p.31). The WWCCR is hosted by the College of Policing (CoP), a professional body for policing in England and Wales which, amongst other things, is committed to identifying, developing and promoting evidence-based policing, which they define as an approach whereby ‘police officers and staff create, review and use the best available evidence to inform and challenge policies, practices and decisions...supported by collaboration with academics and other partners’³. But not all evidence is created equally. In medicine, so often viewed as the exemplar of an evidence-based profession, evidence is conceived of broadly, and is taken to mean any kind of clinically relevant research (Sackett et al. 1996). In a similar vein, the CoP conceive of the ‘best available evidence’ as that which derives from ‘appropriate research methods and sources for the question being asked’⁴. This brings forth the question: what types of evidence are relevant to the needs of crime prevention decision makers?

Enter EMMIE. The EMMIE framework was created in support of the WWCCR as a collaboration between University College London and the CoP. EMMIE is an acronym that

³ <http://whatworks.college.police.uk/About/Pages/What-is-EBP.aspx>

⁴ <http://whatworks.college.police.uk/About/Pages/What-is-EBP.aspx>

encapsulates five broad categories of evidence deemed relevant to crime prevention decision-makers, namely Effect, Mechanism, Moderator, Implementation and Economics (Johnson, Tilley and Bowers, 2015). ‘Effect’ refers to the impact of an intervention. It is the dominant outcome measure in evaluations and systematic reviews in crime prevention. Effect is usually expressed in statistical terms, and in this case represents some likelihood of the expected reduction (or increase) in a crime related outcome. Common formats include odds ratios, standardised mean differences or correlation coefficients (Lipsey and Wilson 2001). The next two elements of EMMIE draw on the scientific realist approach to evaluation (Pawson and Tilley, 1997), which asserts that ‘outcomes unearthed in empirical investigation are intelligible only if we understand the underlying *mechanisms* which give rise to them and the *contexts* which sustain them’ (Pawson and Tilley, 1996, p. 574). ‘Mechanisms’ thus refer to the way in which an intervention is expected to produce the desired outcome(s) (such as reductions in crime). Mechanisms are important because a single intervention can reduce crime in different ways. Take CCTV. Tilley (1993) suggests nine mechanisms through which CCTV installed in car parks might plausibly reduce car crime. These range from deterring prospective offenders by increasing the risk of apprehension to reminding car owners to take necessary precautionary measures (lock doors, close windows and so on) to reduce their risk of victimization. Understanding which mechanism/s is/are responsible for the observed outcome patterns can help strengthen causal attributions (Eck, 2017).

Mechanisms are seldom activated unconditionally, however. ‘Moderators’ refer to the conditions that influence whether an intervention’s mechanisms are triggered and in turn whether the sought-after outcome patterns are produced. In principle, a moderator can include any factor that is causally relevant to the activation of programme mechanisms. In practice, these often relate to the features of the people or places that are the target for intervention, such as the age or gender of recipients or the type of neighbourhood in which an intervention is put

in place. Neighbourhood Watch is a case in point. One mechanism through which Neighbourhood Watch is thought to reduce domestic burglary is by increasing the risk that prospective offenders are apprehended by the police. The logic goes that participation in a Neighbourhood Watch scheme increases the likelihood that affected residents will a) routinely look out for suspicious individuals loitering in their neighbourhood and b) inform the police. The likelihood that residents will perform these behaviours is likely to vary across neighbourhoods, depending upon a number of factors such as trust in the police, community involvement in security partnerships, and willingness to report suspicious behaviours to the police. Failure to notify the police will of course impede the activation of the risk-increase mechanism.

The I of EMMIE denotes ‘Implementation’, referring to the actions or actors necessary to successfully install and maintain an intervention. Implementation failure is widespread in crime prevention (Eckblom, 2010). It is a common reason for projects which have been found to be successful in one setting failing to produce similarly positive results when scaled up or replicated elsewhere (see Tilley, 1996). Finally, E refers to ‘Economics’, which relates to the financial impact of an intervention. This covers several different economic concepts, such as necessary input costs (including variable costs such as staffing and fixed costs such as equipment), direct costs, indirect costs and intangible costs, and calculations which compare the cost of interventions to the benefits gained as a result of any crimes averted (see Manning, Johnson, Tilley, Wong and Vorsina 2016).

A key objective shared by all of the What Works Centres is to make research evidence accessible to decision makers. As David Halpern, national advisor to the What Works network put it, ‘Generating and collating the evidence is of no use if it never reaches the commissioners

and professionals who need it' (2014, p. 6)⁵. In awareness of this, a key output of the WWCCR was the creation of the online Crime Reduction Toolkit⁶ (CRT), organised around the EMMIE framework and designed to summarise and make accessible research evidence for practitioners and policymakers. Two features of the CRT warrant mention at this point. First, the CRT only reports evidence produced by systematic reviews – primary evaluation studies are excluded. This is consistent with other What Works Centres⁷ and speaks to the widely-held assumption that systematic reviews are an authoritative source of evidence (Neyroud, 2018). The second feature concerns the decision by the CoP to present research evidence by *intervention*. Systematic reviews using different units of analysis such as population (i.e. drug-dependent offenders or sex offenders) are excluded.

Figure 1 shows a screenshot of the 'landing page' of the CRT. As indicated above, each entry relates to a crime reduction intervention that has been the subject of a systematic review. It can be seen that for each systematic review there is a symbol relating to the five elements of EMMIE. These symbols serve two purposes. The first is to show whether the review in question reported information on the five dimensions of EMMIE. For example, Figure 1 shows that the systematic review evidence on 'alcohol tax and price policies' contained no information on Implementation and Economics - the relevant symbols in columns four and five are blank. Where evidence relating to the different elements of EMMIE *is* present in a systematic review, the second purpose of these symbols is to give an indication of the *quality* of that evidence, as indicated by the horizontal bars beneath each symbol. This is a topic we shall return to shortly.

⁵ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/378038/What_works_evidence_for_decision_makers.pdf

⁶ See <http://whatworks.college.police.uk/toolkit/Pages/Toolkit.aspx>

⁷ See <https://www.gov.uk/guidance/what-works-network>

[Figure 1 here]

It should be noted that the CRT is not the first nor only effort to distil and present research evidence on the effectiveness of crime reduction interventions. Similar exercises have been undertaken in the US by the National Institute of Justice (see CrimeSolutions.gov), the Centre for Problem Oriented Policing (in the form of problem-specific guides), the Centre for Evidence-based Crime Policy at George Mason University (the evidence-based policing matrix) and in Australia by the University of Queensland (see [Global Policing Database](#); Higgins et al. 2014). The CRT is distinct, however, in reviewing research evidence across a broader range of variables, following the EMMIE framework.

This is deliberate. It speaks to a guiding assumption of EMMIE: that the most useful type of evidence review in the service of crime reduction brings together information on both the effect of an intervention (What Works?) and, inspired by the scientific realist approach, on likely variations in intervention effectiveness (what works for whom, how and under what circumstances?). This is not the standard approach for synthesising research evidence in crime prevention, and is one that requires a mixed-methods approach to systematic review (see e.g. Greene et al 1989), which will be discussed below. It also accords with the notion of Context-Mechanism-Outcome configurations (CMOCs) that underlie the realist approach to evaluation (Pawson and Tilley, 1997). To illustrate in the context of crime prevention, imagine a decision maker contemplating the use of CCTV to deal with a problem of violence associated with the night time economy. For our decision-maker, clearly it is helpful to know what the evidence says about the effectiveness of CCTV as a crime reduction method. However, we contend that it is also useful to know that better results are more likely in some contexts (e.g. car parks) than others (e.g. residential estates), what might account for these variations, what might be done to

maximise the chance of producing the sought-after effects (and mitigating against any adverse effect), and that cost effectiveness varies under different implementation conditions.

A realist synthesis therefore involves casting a wider net in terms of evidence acquisition than is unusually undertaken in a systematic review (which typically only look at evaluation studies of a particular design, such as (quasi)experimental studies), and panning for the ‘golden nuggets’ of information that can help better understand the mechanisms through which an intervention is expected to produce the sought-after effects and the conditions that determine whether said intervention is more or less effective (see Pawson 2006). A truly high-quality review of this ‘mixed method’ type would combine the statistical and methodological standards expected of a meta-analysis undertaken under the banner of the Campbell⁸ or Cochrane⁹ collaboration with an extensive realist review producing ‘middle range theory’ of what is likely to work in what circumstance (Pawson 2006). This is a high bar to reach, and we will argue that the systematic reviews that we have reviewed have seldom achieved this, nor have many others in the crime prevention literature. A movement towards incorporating realist principles in both primary evaluations and evidence syntheses has been seen across the policy landscape and is apparent, for example, in public health research as well as crime reduction (see, for example, Berwick 2008; Creswell 2014; Davidoff 2009; Best et al. 2012; Kastner et al. 2015).

What are the functions of EMMIE?

The central purpose of EMMIE is to act as a framework to effectively communicate research evidence to those with a responsibility for and competency to deal with crime. More generally, it aims to foster a broader conception of what constitutes relevant research evidence, moving

⁸ See <https://www.campbellcollaboration.org/>

⁹ See <http://www.cochranelibrary.com/>

beyond just ‘what works?’ (Tilley and Laycock, 2017). To this aim, presently EMMIE has been used in two main ways. The first is as a tool to appraise evidence in systematic reviews in crime reduction. The EMMIE appraisal tool contains, at the time of writing, 105 dimensions of assessment corresponding to the five dimensions of EMMIE (see Tompson et al, 2015¹⁰). Items were developed following an assessment of existing evidence appraisal instruments (such as the Cochrane risk of bias tool¹¹), reporting guidelines for systematic reviews (such as PRISMA¹²) and through discussion with researchers experienced in crime reduction evaluation. The 105 items fall into two broad categories, those that relate to the evidence *reported* in systematic reviews (What was the sample size? What data were used? Did crime go up or down?) and those that relate to the methodological *quality* of that evidence (see Johnson, Tilley and Bowers, 2015).

To illustrate how the tool is used and to distinguish between these two classes of item, consider Effect. For this dimension of EMMIE, there are ten items on the evidence reported in a systematic review. These items relate to details of the review being assessed such as the ‘type of effect size’, ‘number of studies used in meta-analysis’, ‘number of participants across action groups’ and so on. There are also fourteen items concerned with the methodological quality of evidence relating to the Effect of the intervention under review. These items are framed as a series of ‘yes/no’ questions: ‘is there a transparent and well-designed search strategy?’ and ‘is there an assessment of the influence of study design?’. Unlike the previous set of items in which information is simply extracted from a systematic review, items concerning the quality of evidence require judgments to be made as to whether, for example, the reported search strategy is deemed to be ‘transparent and well-designed’. A codebook was designed collaboratively to

¹⁰ Available to be downloaded at: <http://discovery.ucl.ac.uk/1462093/>

¹¹ See http://handbook.cochrane.org/chapter_8/8_assessing_risk_of_bias_in_included_studies.htm

¹² See <http://www.prisma-statement.org/> and Moher et al 2009

systematise some of these judgements and is embedded within the appraisal tool. The responses to each of these fourteen items are then combined to produce an overall evidence quality score per systematic review (a formula within the tool provides guidance for this score, but this can be overwritten by the coder/s).

There are similar items for the four remaining dimensions of EMMIE, with each dimension giving rise to an overall evidence quality rating per systematic review (for more details see Tompson et al 2015). As expressed above, this quality profile relates solely to the methodological quality of the evidence, not the evidence itself (i.e. whether an intervention was found to be effective). This overall quality rating is presented as a five-point scale (see Figure 2, with 0 denoting the lowest quality of evidence and 4 denoting the highest quality of evidence). It follows that across the five elements of EMMIE, a systematic review can score a maximum of 20 points. As will become clear, many of the systematic reviews identified as part of the WWCCR fell some way short of top marks, for reasons we will speculate upon later.

[Figure 2 here]

So far we have described EMMIE as a tool for assessing the evidence reported in *existing* systematic reviews. A second way EMMIE has been used to date is to support the design and conduct of systematic reviews, by encouraging reviewers to search for and synthesise evidence on each of the five elements of EMMIE. Used this way, it is plausible that EMMIE might also be put to more general use to support the conduct and reporting of primary evaluation studies. This is yet to happen. However, as will be discussed shortly, efforts undertaken as part of the WWCCR to systematically review the literature on alley gating, security tags and electronic monitoring of offenders, respectively, revealed knowledge gaps pertaining to several of the

elements of EMMIE in those primary studies selected for review.

These are not the only possible uses of EMMIE. Fleming, Fyfe and Wingrove (2016) describe how EMMIE has been used as a framework to introduce practitioners to the notion and practice of evidence-based policing. Bowers et al. (2017) discuss how the EMMIE appraisal tool might be adapted for use as reporting guidelines for systematic reviews in crime reduction. Moreover, whilst so far applied only to studies in crime reduction, Tilley (2016) argues that EMMIE might usefully inform the collection and assessment of evaluation evidence in cognate fields. To this aim, Bowers et al. (forthcoming) recently drew on EMMIE to appraise the evidence-base as it relates to terrorism studies, cybercrime and the forensic sciences.

In what follows, we limit our discussion to the two main ways in which EMMIE has hitherto been used, and seek to answer the following questions: How do existing systematic reviews in crime reduction fare according to the EMMIE framework? And, how and with what result was EMMIE used to inform three new systematic reviews undertaken as part of the WWCCR?

What does EMMIE tell us about systematic review evidence for crime prevention?

One of the first tasks of the research programme undertaken in support of the WWCCR was the systematic mapping of the evidence base for crime reduction. This was done in accordance with the principles of systematic searching that characterize systematic reviews (Tompson and Belur, 2015), with a twofold inclusion criteria that specified a study had to a) be a systematic review or meta-analysis, and b) have a measurable crime reduction outcome (not an intermediate outcome such as truancy that might plausibly be linked to crime but which is not illegal).

The strategy and results of this literature search have been reported in detail elsewhere (see Bowers et al. 2014). Briefly, 337 evidence syntheses made it through the two-stage screening process. A breakdown of the 337 reviews is provided by Tompson and Thornton (2016). This further delineated reviews into those that focused on *single* interventions and those that included *multiple* interventions. This distinction is important for making causal attributions about the effectiveness of a given intervention. Where multiple interventions have been put in place, it is often more challenging to isolate the effect of a specific intervention (and, potentially, a specific mechanism). The definition of a single intervention is somewhat nebulous, however, and sometimes it is tricky to establish the fidelity of interventions across studies which share the same name unless an intervention is patented, or delivered within a very rigid set of guidelines. Often it cannot be considered to be the same intervention across different implementation teams or contexts. We therefore took a pragmatic approach and looked for interventions that shared a similar rationale and/or implementation conditions, but acknowledged that this was still a subjective judgement. Using this criterion, the 337 reviews originally identified were reduced to 70 reviews on 44 intervention topics, published between 1995 and 2015.

Each of the systematic reviews that met our inclusion criteria were then subject to evidence appraisal using the EMMIE coding instrument described earlier. At least two coders appraised each review, and a moderation session was held to reconcile any differences that arose. The principal coder subsequently wrote up the evidence synthesis into a narrative that now appears on the CRT. In what follows, we discuss the general trends that emerged from using the EMMIE-Q (quality of information) codes to rate and rank the interventions featured in these 70 reviews. These are summarised in Table 1.

[Table 1 here]

Taking each of the EMMIE dimensions in turn, we see that across the 70 reviews the quality score for effect has a median of 3 and a mean of 2.5. As Figure 2 indicates, a full score of 4 represents a review that has been attentive to at least five key sources of bias in a meta-analysis (such as selection bias, statistical conclusion validity etc). Tompson et al. (in preparation) show that the effect score in these 70 systematic reviews has improved over time, and speculate that this may reflect the influence and guidance promulgated by organisations such as the Campbell Collaboration.

With respect to the mechanism dimension of EMMIE, a handful of syntheses did not report any information relating to the presumed mechanisms thought to be associated with the intervention under review (n=9, 13%). Furthermore, few evidence syntheses were found to provide more than a blanket statement of how the intervention is assumed to produce its effects, which was scored a 1 on the EMMIE rating scale (n=36, 51%). A common statement was that an intervention ‘deterred’ offending, but little was said as to how this deterrence might be achieved. If a review explicated what was meant by deterrence (i.e. specific or general) and/or discussed the means by which deterrence might be achieved (e.g. by making it more likely that the offender would be apprehended), this was counted as a detailed description of the theory of change (see Figure 2), which was the case in around a fifth of the reviews (n = 15, 21%), albeit with many having a different mechanism than deterrence specified. A full theoretical model with predictions of outcome patterns was less commonly reported (n = 8, 11%) and an empirical test of the putative mechanisms was observed in only two reviews (on imprisonment and drug courts). This distribution is reflected in the median, mean and standard deviations presented in Table 1. As mentioned previously, it is our contention that understanding how an intervention produced the observed outcome patterns is important, particularly when considering whether the same intervention stands a good chance of producing positive results

elsewhere. It was therefore disappointing that so few evidence syntheses paid adequate attention to how the interventions were assumed to work.

Information on possible moderators of intervention effectiveness was more common than that concerning intervention mechanisms. This may be attributable to moderators occupying a more familiar place in systematic review terminology and methods textbooks (Gough, Oliver and Thomas 2013). As shown in Figure 2, our coding system strived to tease out those reviews which appeared to analyse moderator variables that were conveniently at hand (*post-hoc*) and those that considered the mechanism when searching for and analysing moderators (*a priori*), with the latter achieving a higher quality score. A common example here is reviews which theorise *a priori* that gender differences will have an impact upon the results, and provide an explanation as to why, compared to those reviews which, say, use geographical differences (e.g. countries) as a convenient post-hoc moderator that emerged from the primary studies. Distinguishing between the two types of approaches to moderator analysis was not always straightforward. In particular, older systematic reviews did not have a common reporting format, and some public health journal papers included in our final sample were shorter than is common in the social sciences and so it was not always possible to discern if moderators had been selected before or after the primary studies were analysed. Our approach to this was to scrutinise the introductory information for evidence of theoretically motivated moderators. For example, if a review reported that young offenders had been found to be more amenable to a particular form of treatment than older offenders, and age was later used in moderators, we counted this as theoretically motivated. Clearly, our judgement of the type of moderator analysis undertaken might be influenced by reporting bias, insofar that reviews that contained more information that we were looking for, and were written clearly, were likely to score a higher mark.

From Table 1 we see that post-hoc analysis of moderator variables is common, with the median and mean value indicating that most reviews had a score of 1 or 2 ($n = 21, 30\%$ and $29, 41\%$ respectively). It was less common for reviews to have theoretically-grounded moderator variables that they actively sought from primary studies.

The descriptive statistics for Implementation were similar to those for Mechanism, with a general tendency for limited information to be reported. To score a 1 on this dimension, reviews had to make ad-hoc comments on implementation, albeit the 30 (43%) that did so were very brief and often just covered some well-known barriers (e.g. staffing issues). To score a 2, reviews had to make concerted efforts to document implementation challenges or enablers (16 reviews did so, 23%), which might take the form of discussing resourcing (e.g. how many staff, how much training is required) or the structure of a therapeutic component. It was rare for a review to report an evidence-based account of implementation conditions and/or analysis which illustrated adherence to programme fidelity. Those that did so tended to not just focus on the barriers to implementation – for example, participation rates – but also ways to overcome them.

The scores for the economics section indicated a dearth of economic data in the analysed reviews. Only six reviews (9%) attempted any kind of synthesis across the primary studies, with five of these solely estimating direct costs (and/or benefits). A review on alley gates (Sidebottom et al. 2017a) estimated both direct and indirect costs (and/or benefits), thus scoring 2 out of a possible 4 on this dimension. The economics dimension is probably the clearest illustration of the difficulties of appraising the quality of evidence at the evidence synthesis level. Monetised figures for the costs or benefits of an intervention are rarely comparable across studies, due to a multitude of reasons but perhaps most notably owing to the discretion with which decisions are made on what should or should not be included in costings (Manning et al

2016). Systematic review authors may be disinclined to attempt to harmonise and synthesise these due to the analytic difficulties involved (Manning et al 2016). In light of these challenges, we recognise that it is perhaps unfair to give equal weighting to Economics as we do other elements on the EMMIE rating scale. Simply put, given the poor quality of economic data contained in primary evaluations, reviews will usually score poorly on the EMMIE quality appraisal scale.

To summarise, based on an assessment of 70 systematic reviews of single interventions that contained a quantitative crime reduction outcome measure, the effect dimension of the evidence base in crime reduction appears to be in good health, and improving over time. It is however worth noting that due to the inclusion criteria employed in this exercise – must have a measurable crime reduction outcome – we are likely privileging effect over the other dimensions of EMMIE in the search for relevant systematic reviews, and it is thus to be expected that this dimension receives the best overall quality scores. The mechanisms, moderators, implementation and economics dimensions are less well-developed in the crime reduction literature we have assessed. Just as methodological issues that have compromised the reliability of meta-analysis have been made prominent in guidance on systematic review methodology (see, for example Gough, Oliver and Thomas 2013) for effect, in our view, what now needs to happen is a similar elevation of awareness of the issues surrounding mechanisms, and those factors which support the mechanism firing. Our hope is that the EMMIE framework will contribute towards this goal.

EMMIE inspired systematic reviews: results and observations

Another strand of work undertaken in support of the WWCCR was the production of twelve novel systematic reviews of crime reduction initiatives. The topics of these reviews were

selected through consultation with stakeholders and the CoP¹³. In this section, we focus on three reviews led by staff at UCL on the topics of alley gating (Sidebottom et al 2017a), security tagging to prevent retail theft (Sidebottom et al 2017b) and the electronic monitoring of offenders (Belur et al, 2018). All three reviews focus on situational interventions with a common aim of reducing (re)offending through securing a place, an item or a person. The reviews were undertaken with the objective of collecting and synthesising evidence on each of the five elements of EMMIE, and were conducted over a period of three years, by essentially the same team, with lessons learned from one review feeding into the next. The most important of those lessons will be the focus of this section, including both methodological and practical issues encountered.

Conducting these three reviews brought home the resource-intensive nature of systematic reviews more generally (see Borah, Brown, Capers and Kaiser 2016), but especially one that aims to combine meta-analysis with realist synthesis. It requires a team of researchers with adequate methodological skills to conduct a systematic review and also what is essentially mixed methods research. The detailed coding tools used and the difficulty in distinguishing between the non-effect elements of EMMIE meant that two researchers had to blind code each study and then reconcile their coding. While disagreements were rare, any that did occur had to be resolved by a third researcher who was also familiar with the coding process.

Our experience raised fundamental questions about the replicability of a mixed-methods EMMIE-inspired review given the considerable amount of judgement required at various stages of the review – in the selection of studies (see Belur et al, 2018), and during the coding phase. Studies were initially selected for appraisal in the conventional way when conducting a

¹³ See: http://whatworks.college.police.uk/Research/Systematic_Review_Series/Pages/default.aspx

systematic review – through a combination of electronic databases, searches of the grey literature and consultation with experts in the field. Inter-rater reliability tests were conducted to ensure as much validity in study selection as possible, yet for those studies which were judged eligible for a realist syntheses for elements of EMMIE other than Effect, a level of subjectivity remained. It is arguably easier and more unequivocal to select studies on the basis of, say, research design than it is study content. A meta-analysis typically does the former and can therefore generally ensure a high degree of consistency and replicability. A realist review generally does not exclude studies based on research design and therefore judgements on eligibility were made on the content of each article identified in our searches. For example, to be included in the realist arm of the systematic review of alley gating, articles had to report ‘substantive information’ on mechanisms, moderators, implementation or economics as it relates to alley gates, irrespective of the type of study design. Determining eligibility based on content, in our experience, was difficult to mechanise and hence time consuming.

Given the two strands of the EMMIE-inspired systematic review we could either have two entirely separate inclusion criteria and two separate sets of studies, as in the case of the alley gating and tagging reviews, or only one set of studies suitable for meta-analysis which were then mined for the MMIE elements of the review, as was in the case of the electronic monitoring review. For the latter review, the specific inclusion criteria we used focused heavily on reporting of quantitative effect sizes for crime prevention, which necessarily excluded some information which may be of use to realist synthesis or aspects of the intervention that were aimed at achieving outcomes other than crime reduction (such as reducing the prison population, rehabilitation or reducing costs of imprisonment). The choice of one or the other strategy dictating the inclusion criteria was influenced by the number and quality of primary evaluations once they had been retrieved. It placed a greater burden of responsibility on researchers to make judgements about eligibility of studies as well as deciding what was an

‘acceptable’ number of studies and what information was ‘good enough’ to be included¹⁴. We also had to make decisions about why certain studies had to be excluded because they did not contain sufficient ‘substantive’ information in order to keep the number of studies included manageable in terms of time and resources available. Often, specifying the exact criteria on the basis of which decisions were made about what could be considered ‘substantive’ was challenging and subjective.

Once studies had been chosen for inclusion, they needed to be coded with each EMMIE element represented by multiple variables. Even though each study was double blind coded and then reconciled, inevitably overall decision-making in the coding stage was subjective. Although within the team there was a fair amount of consistency in the application of codes, we recognise that judgement calls were often based on collective subjective understanding that might not be universally shared. Furthermore, given the complex nature of coding MMI and E in particular, it was very difficult to conduct tests of inter-rater reliability which are also associated with the most methodologically-stringent systematic reviews. Any attempt at trying to calculate even basic percentage agreement for each coding category would have been very difficult. Coding open-ended data, especially dealing with large number of codes, makes coding decisions very complex and can often require several rounds of coding (and reconciliation exercises) to achieve respectable inter-rater agreement (Hruschka et al 2004).

Discussions amongst members of the team revealed some uncertainty around whether we were trying to artificially fit complex constructs into specific boxes due to the systematic review process itself. When attempting to combine the realist elements of EMMIE with a meta-analysis, we fit neither fully into one box nor the other. We cannot consider any of the

¹⁴ The challenge of knowing when to stop searching for information is a challenge to all realist reviews, and something which Pawson (2004) discusses in terms of reaching saturation of knowledge.

systematic reviews conducted within the remit of the WWCCR to be fully ‘realist’ in nature, and refer to them instead as realist-inspired (Sidebottom et al 2017a; 2017b). We try to draw upon realist principles to ensure that important information other than just the effect size is included within our results, but the very nature of systematic reviews means that it must be transparent and replicable, something which a realist review may not be. It appears then, that there is a tension between the realist approach and the systematic review approach to evidence synthesis, and this can strain the reliability and/or value with which the eventual evidence is seen by readers (Gough, Oliver and Thomas 2013, Petticrew and Roberts 2008).

Thus limitations posed by resources and time, and our focus on the crime prevention aspect of interventions meant our search had to necessarily pivot around the Effect element, possibly to the detriment of the four MMIE elements in more qualitative or narrative studies. Despite that, overall, we feel that our experience over the three reviews has led to a workable compromise which allows some of the ‘best’ evidence from both effect and other elements to be presented. We suggest that EMMIE is a useful addition to the body of work that has tried to provide standards for realist oriented work (for example the RAMESES project¹⁵) while at the same time combining it with meta-analysis (Petticrew and Roberts 2008).

We now turn to the results of our systematic reviews regarding the search for information for each EMMIE element. Our first observation is that the majority of crime prevention primary studies that we encountered in the process of undertaking these three reviews did not report their findings using EMMIE terminology. While many studies included a literature review which alluded to the mechanisms through which the intervention was thought to give rise to reductions in crime, these were often not explicitly stated. One common example concerns the

¹⁵ See <http://www.ramesesproject.org/>

primary studies on the electronic monitoring of offenders which referred to the intervention increasing tagged offenders' levels of self-control, without discussing how changes in levels of self-control might be associated with changes in (re)offending. Primary studies which go to the extent of creating full theoretical models are few and far between. This left us trying to tease out this information, mainly through inference and presumption. It followed that the investigation of intervention mechanisms, in terms of whether they were activated and, if so, whether they were responsible for the observed outcome patterns, was even rarer, especially since this required an in-depth discussion of the link between mechanisms, intervening outcomes and the final crime prevention effect of the intervention.

Although the majority of the primary studies consulted as part of these three reviews did contain some explanation for the links between the intervention and the outcomes observed, this was, unfortunately, not made clear or explicit enough in many cases to allow the reader to understand the mechanisms thought to be at play and how they could be manipulated to test their impact upon crime reduction. In our alley gating review, for example, we were able to construct a number of context-mechanism-outcome configurations which illustrated the links between intervention, mechanism and context and the outcome patterns expected (see Figure 3 below). Both the reviews on security tagging and electronic monitoring included logic models which visually represent the linkages between the mechanisms, moderators, implementation factors and outcomes. These were designed to guide future primary evaluations to be able to identify intermediate outcomes and collect data *a priori* to test these mechanisms.

[Figure 3 here]

The information we were able to collect on moderators highlights a different issue. Most primary studies consulted in our three systematic reviews included the most convenient or

easily accessible moderator-based analysis e.g. male vs female offenders, adults vs juveniles, USA vs European contexts etc. Nevertheless, the paucity of explanation behind *why* such contextual differences should affect the outcomes continues. In many cases it was just presumed that the reader will understand why, for example, adults and juveniles should benefit from this intervention differently, or why it should have a greater impact upon males than females.

Fortunately, the three situational measures which formed the basis of the UCL systematic reviews discussed implementation factors in great detail, allowing this part of the review to provide academics and practitioners with knowledge on ways to try to improve the successful implementation of these interventions. However, as seen in the section above, this is not ordinarily the case. With regard to electronic monitoring and tagging, both of these interventions require complex systems in order to be effective. Developing an understanding of the elements of these systems and the way in which they can be altered is necessary to ensure their successful implementation. For example, in order for tagging of goods to be effective, the technology should work, the shop staff should be alert, trained, and ready to respond to an alarm, and there should be willingness to prosecute or take suitable action to reinforce the message that shoplifting will carry negative consequences.

The information regarding economics found within the three reviews we conducted mostly mirrored findings from the wider evidence base synthesised in Table 1. However, primary studies included for all three reviews contained some economic information, even though we were only able to actually synthesise it for the alley gating review. The fact that much of the evidence base on alley gating came from Home Office funded initiatives meant that the final reports to the Home Office contained a wealth of information about the costs of the scheme. This allowed us to conduct a cost-benefit analysis of alley gating (Sidebottom et al 2017a). The

information related to tagging in retail environments was sparser, and those primary evaluations with the richest potential evidence were retailer reports which were confidential. The complex nature of the system surrounding electronic monitoring meant that the economic information we encountered on this topic was varied, with costings about different parts of the system sometimes including a breakdown in terms of equipment, staffing and incidental costs in some studies, but not all. However, the data overall was not suitable for a statistical synthesis or cost-benefit analysis.

Our experience of reported data embedded in the primary evaluations along EMMIE dimensions indicated that for some elements, such as mechanisms and economics, distinguishing between and coding information relating to different categories was fairly straightforward. For other elements of EMMIE, especially moderators and implementation, the distinction often seemed ambiguous. For example, untangling whether a particular aspect was a moderator (i.e. pre-existing and unalterable given condition) or part of implementation (something that could be altered to the boost effectiveness (or otherwise) of the intervention) was often difficult. For example, while conducting the systematic review on tagging, an examination of the shop layout itself could be considered as both Moderator and Implementation – depending on whether the size of the store and its layout is considered to be pre-existing (moderator) or something than can be altered (implementation). This is unsurprising. Many of the activities undertaken in the name of implementation might plausibly influence the activation of intervention mechanisms, the definition of a moderator used in EMMIE. By contrast, many of the factors that might affect the activation of causal mechanisms have little to do with implementation. In practice, however, when coding information from primary studies as part of the systematic review, coders dealt with such grey areas through in-depth discussion and the acceptance that some form of agreement at the coding stage was enough as long as pertinent information was being recorded – better to include the information

for consideration by decision makers than omit owing to disagreement as to whether it relates to Moderator or Implementation.

Further, our experience indicated that often the same element figures in the EMMIE framework both as moderator and implementation and often triggers different mechanisms depending upon the particular conditions. A clear example of this was found in our systematic review of electronic monitoring of offenders. In this case, the type of technology used (Radio Frequency (RF) or GPS) was a moderator as often it was predetermined and fixed for every programme. However, its effectiveness and how it operated depended upon a set of implementation issues such as access to electricity (for RF), satellite network (for GPS) and suitable communication channels between various agencies involved. Further, the type of technology chosen affects the particular mechanism it relies upon in order to be effective. In other words, the moderator and its implementation in turn affects the mechanism activated in order for electronic monitoring to reduce reoffending. An example of this would be the use of active GPS technology which tracks movement 24 hours a day and may be perceived as leading to a greater risk of getting caught offending. In contrast, RF technology often requires the offender to be confined to one location during a curfew (often overnight) which suggests a more preventive mechanism. Thus, GPS can be posited to increase the risk (of getting caught), whereas RF could be said to reduce the opportunity to commit crimes, that would otherwise be committed during curfew hours. But both these mechanisms would be activated provided the implementation were appropriate, i.e. the technology works, there is good communication and data sharing between monitoring and supervising agencies and the response to a breach is swift and certain.

What next?

This paper set out to describe the EMMIE framework which grew out of the WWCCR, discuss the ways in which EMMIE has been used to date, and present observations and critical reflections from conducting systematic reviews aligned to the five elements of EMMIE. Central themes of this paper include: (1) the observation that much of the evidence reported in the crime reduction literature is limited to that which speaks to the effectiveness of interventions; there is a noticeable evidence drought in terms of those factors that might plausibly give rise to variations in the likely effect of crime reduction interventions and (2) that locating and integrating different types of research evidence that would help alleviate this drought is a complex and time-consuming task, albeit one which we contend is helpful to produce the sort of transformative change desired by advocates of evidence-informed crime reduction.

We have noted that while systematic reviews of (single) crime prevention interventions score highly on the effect dimension of EMMIE, and exhibit an upward trajectory over time, scores on the mechanism, moderator, implementation and economic dimensions are less well-developed in the systematic reviews we assessed. One reason that past studies might have said little on the non-effect dimensions of EMMIE relates to difficulties synthesising all the information that is unearthed. In many cases dissimilar, or even conflicting information is presented across primary studies but there is often little attempt by review authors to tease out the implementation or contextual factors that might cause the contradictory results. In other words, interpretation of the integrated evidence might either be done with little consideration of the validity of doing so, or not attempted at all. It is important to note, of course, that primary studies may indeed not report this information, even if it was available to them. This may be due to issues such as journal word limits which often work against reviews which would like to report such information.

It bears further discussion whether giving equal weighting to each element of EMMIE while quality appraising, as the current rating scale does, is indeed fair. This is for two reasons: firstly, all dimensions of EMMIE rarely get equal coverage in peer reviewed papers. Secondly, often data is missing or incomplete to facilitate analysis on all EMMIE dimensions. Our research indicated that mechanisms are rarely tested except post-hoc and often economic information and analysis is missing or incomplete. There is an argument for the current version of the weighting of the EMMIE rating scale to be revised to fairly reflect the quality of the wide spectrum of evaluation research that exists.

A further key lesson that has emerged from doing this work is that appraising evidence at the *evidence synthesis* level requires expertise in both evaluation theory and systematic review methodology. Deciding upon an inclusion criteria for evidence across all elements of EMMIE is challenging as the types of publication or grey literature which contain this information varies greatly. We found that each review presents unique challenges to the quality appraisal tool, and many coding decisions needed to be taken in consultation with the team. There is an elongated learning curve for novice coders to conquer before they feel confident in making the judgements required by the coding tool. This emphasises the importance of having robust quality assurance procedures in place when undertaking such a task. The resources required to conduct an EMMIE-inspired review are necessarily much greater than a traditional meta-analysis which tends to look for only one of the EMMIE elements (typically effect).

It is heartening to see that the original exposition of EMMIE (Johnson, Tilley and Bowers 2015) and the systematic review on alley gating (Sidebottom et al 2017a) have been published in leading journals in the field (Journal of Experimental Criminology and Justice Quarterly respectively). This suggests an openness to, at least amongst academics, the use of EMMIE and the evidence which an EMMIE-inspired review presents. It remains to be seen whether

EMMIE garners the same level of interest among practitioner audiences. In terms of practitioner engagement, to date the evidence indicates a mixed reception of some of the ideas inherent in the EMMIE framework (see Hunter, May and Hough 2017 for a more extended commentary on this). It is clear that the CRT has elicited mixed reactions from practitioners and policymakers. It appears that those more familiar with the concept of evidence-based policing (often more senior ranking staff) often support the principles of the CRT, but equally, some found the more extended framework rather academic and frustrating (ibid). Nevertheless, there are reasons to be optimistic that such ideas are gaining momentum. For example, the realist approach, and EMMIE more specifically, is now an integrated part of the training for Police Now officers¹⁶. However, to truly move forward it is important that thought is given to a number of issues; these are- how to resource the gathering of evidence, how to provide support in adequate documentation of contextual aspects of interventions and how to encourage crime preventers to own these ideas and integrate them in their day-to-day practices.

Finally, a specific recommendation concerns encouraging evaluators to consider EMMIE in the design and conduct of primary evaluations. As we have seen in our reviews of the alley gating, tagging and electronic monitoring literature, there is wide variation in the extent to which primary studies report evidence on the different elements of EMMIE. We suspect similar patterns may be true for other form of crime reduction measures. It is our hope that greater use of the CRT, structured as it is around EMMIE, will serve not only to inform decisions about how best to deal with presenting crime and community safety problems but also influence the way in which evaluations of crime reduction are commissioned and/or conducted. More specifically, by encouraging evaluators to think through how they might collect data that speaks to the different elements of EMMIE, beyond just impact on crime.

¹⁶ For more information about Police Now, see <https://www.policenow.org.uk>

This applies to both researchers and practitioners. In terms of the latter, there are several ways in which we might further embed EMMIE thinking. This includes the facilitation of police-academic research partnerships as well as training in the sorts of research methods that might plausibly be drawn on to collect relevant information on the different elements of EMMIE. Ultimately, however, we suspect that a tipping point denoting the widespread use of EMMIE, should it happen, will depend on reliable evidence demonstrating the added crime reduction benefits brought about by using EMMIE. Evidence of this kind is presently lacking, owing to the recent creation of the EMMIE framework. Consequently, it is important that advocates for EMMIE subject it to the same sort of empirical scrutiny called for in evaluations of the crime reduction measures, namely, an assessment of how it is applied in practice and whether its use is associated with the sought-after outcomes.

Indeed, there is a risk of conveying to the reader that EMMIE is the solution to all issues with the crime reduction evidence base. Frameworks that use categorisation and quantitative scoring have the inherent problem that they lack flexibility and force decisions when evidence doesn't fit neatly. This is evident from the inter-rater reliability statistics which in a perfect situation would demonstrate complete agreement. We have already acknowledged the issue with the over-simplistic weighting of each of the elements of EMMIE as equal. Further to this we put high demands on evaluators when it comes to undertaking and reporting their research. Our experience with primary EMMIE compliant evaluations revealed that it is a time-consuming practice and is additionally a significant challenge to fit in all the components in reporting the results. This is particularly true in the context of academic journal contributions that are often restricted in terms of word count. In many cases elements of EMMIE are reported across multiple papers for a single evaluation which adds to the workload associated with consolidation. Finally, it has been criticised by some for being too academic in tone and

difficult to digest quickly in the busy world of practice. These limitations of the EMMIE framework set an agenda for future refining and validation processes. In the meantime, we suggest that there are many advantages to the holistic approach that the framework offers and that adoption of the approaches described here should lead to better informed crime prevention in practice.

References

- Belur, J., Simon, M., Thornton, A. and Tompson, L. (2018) 'Inter-rater reliability testing in systematic reviews', *Sociological Methods and Research*, forthcoming
- Belur, J., Thornton, A., Tompson, L., Manning, M., Sidebottom, A., Johnson, S. and Bowers, K. (2018) 'A Systematic Review of the Effectiveness of the Electronic Monitoring of Offenders', in preparation
- Berwick, D. (2008) 'The science of improvement', *Journal of the American Medical Association*, 299, 1182-1184
- Best, A., Greenhalgh, T., Lewis, S., Saul, J., Carroll, S. and Bitz, J. (2012) 'Large system transformation in health care: a realist review', *The Millbank Quarterly*, 90(3): 421-456
- Borah, R., Brown, A., Capers, P. and Kaiser, K. (2017) 'Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry', *BMJ Open*; 7:e012545.; doi:10.1136/bmjopen-2016012545
- Bowers, K., Gill, P., Morgan, R., Meiklejohn, S. and Johnson, S.D. (2017) 'Evidence 'Challenges' in the Security and Crime Sciences', In Farrell, G. and Sidebottom, A. (eds), *Realist Evaluation for Crime Science: Essays in Honour of Nick Tilley*, Routledge
- Bowers, K., Tompson, L., Sidebottom, A., Bullock, K. and Johnson, S. D. (2017) 'Reviewing Evidence for Evidence-Based Policing', In Knutsson, J. and Tompson, L. (eds) *Advances in Evidence Based Policing*, London: Routledge, forthcoming
- Bowers, K., Tompson, L. and Johnson, S.D. (2014) 'Implementing information science in policing: mapping the evidence base', *Policing*, 8:4, 339–352

Clarke, R. (2017) 'The Links between Realistic Evaluation and Situational Crime Prevention', In Farrell, G. and Sidebottom, A. (eds), *Realist Evaluation for Crime Science: Essays in Honour of Nick Tilley*, Routledge.

Creswell, J. W. (2014) *Research design: Qualitative, quantitative, and mixed methods approaches*, Thousand Oaks, California: SAGE Publications

Davidoff, F. (2009) 'Heterogeneity is not always noise: lessons from improvement', *Journal of the American Medical Association*, 302(23), 2580-2586.

Eck, J. E. (2017) 'Evaluation for lesson learning', In Tilley, N. and Sidebottom, A. (eds) *Handbook of crime prevention and community safety*, 2nd edition, Routledge

Eckblom, P. (2010) *Crime prevention, security and community safety using the 5Is framework*, Springer

Fleming, J., Fyfe, N., & Wingrove, J. (2016) *Evidence-informed Policing: The Design of a Pilot Training Programme*. Report produced for the College of Policing. Gough, D., Oliver, S. and Thomas, J. (Eds.) (2013) *An introduction to systematic reviews*, London: Sage Publications Ltd.

Fleming, J. and Wingrove, J. (2017) "'We Would If We Could ... but Not Sure If We Can': Implementing Evidence-Based Practice: The Evidence-Based Practice Agenda in the UK', *Policing: A Journal of Policy and Practice*, 11:2, 202–213.

Gough D, Maidment C, Sharples J (2018). *UK What Works Centres: Aims, methods and contexts*. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London. ISBN: 978-1-911605-03-4. Available from <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3731>

Greene, J. C., Caracelli, V.J and Graham, W.F. (1989) 'Toward a Conceptual Framework for Mixed-Method Evaluation Designs', *Educational evaluation and policy analysis*, 11; 255

Hruschka, D. J., Schwartz, D., St. John, D. C., Picone-Decaro, E., Jenkins, R. A. and Carey, J. W. (2004) 'Reliability in coding open-ended data: Lessons learned from HIV behavioral research', *Field methods*, 16(3), 307-331

Hunter, G., May, T. and Hough, M. (2017) An Evaluation of the 'What Works Centre for Crime Reduction' Final Report, available at www.icpr.org.uk/media/43599/icpr_final_evaluation_wwccr.pdf

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*, 11(3), 459-473

Kastner, M., Perrier, L., Hamid, J., Tricco, A. C., Cardoso, R., Ivers, N. M., and Graves, L. (2015) 'Effectiveness of knowledge translation tools addressing multiple high-burden chronic diseases affecting older adults: protocol for a systematic review alongside a realist review', *BMJ open*, 5(2), e007640

Knutsson, J. and Tompson, L. (2017) (Eds.) *Advances in evidence-based policing*, Abingdon, UK: Routledge.

Lipsey, M. W. and Wilson, D. B. (2001) *Practical meta-analysis*, Thousand Oaks, CA: Sage Publications.

Lumsden, K. and Goode, J. (2016) 'Policing Research and the Rise of the 'Evidence-Base: Police Officer and Staff Understandings of Research, its Implementation and 'What Works'', *Sociology*, <http://doi.org/10.1177/0038038516664684>

Manning, M., Johnson, S. D., Tilley, N., Wong, G. T., and Vorsina, M. (2016) *Economic Analysis and Efficiency in Policing, Criminal Justice and Crime Reduction: What Works?*, Basingstoke: Palgrave Macmillan

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. and Prisma Group. (2009) 'Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement', *PLoS med*, 6(7), e1000097.

Pawson, R. (2002) 'Evidence-based policy: The promise of 'realist synthesis'', *Evaluation* 8(3) 340-358

Pawson, R. (2006) *Evidence-Based Policy. A Realist Perspective*, London: Sage

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

Petticrew, M. and Roberts, H. (2008) *Systematic reviews in the social sciences: A practical guide*, John Wiley & Sons

Sackett D.L., Rosenburg W.M.C., Gray J.A.M., Haynes R.B. & Richardson W.S. (1996) Evidence-based medicine: what it is and what it isn't. *British Medical Journal* 312, 71–72.

Sherman, L. (1998) Evidence-Based Policing. Ideas in American Policing Series, Washington, DC: Police Foundation, Department of Justice, Office of Justice Programs, National Institute of Justice.

Sherman, L. (2013) 'The rise of evidence-based policing: Targeting, testing, and tracking', *Crime and justice*, 42(1), 377-451.

Sidebottom, A., Tompson, L., Thornton, A., Bullock, K., Tilley, N., Bowers, K. and Johnson, S. D. (2017a) 'Gating Alleys to Reduce Crime: A Meta-Analysis and Realist Synthesis', *Justice Quarterly*. DOI:10.1080/07418825.2017.1293135

Sidebottom, A., Thornton, A., Tompson, L., Belur, J., Tilley, N., Bowers, K. and Johnson, S. D. (2017b) 'A Systematic Review of Tagging as a Method to Reduce Theft in Retail Environments', *Crime Science*, in press

Tilley, N. (1996) 'Demonstration, exemplification, duplication and replication in evaluation research', *Evaluation*, 2(1), 35-50

Tilley, N. (2016) 'EMMIE and engineering: what works as evidence to improve decisions?', *Evaluation*, 22(3) 304-322

Tilley, N. and Laycock, G. (2017) 'The why, what, when and how of evidence-based policing', In Knutsson, J. and Tompson, L. (eds), *Advances in Evidence-Based Policing*, Routledge

Tompson, L. and Thornton, A. (2016) 'Mapping the evidence', Presented at a workshop for Police Scotland. Edinburgh, 2nd February. Available online: <https://www.slideshare.net/LisaTompson/ww-scotland-mapping-the-evidence-58109443>

Tompson, L. and Belur, J. (2015) 'Information retrieval in systematic reviews: a case study of the crime prevention literature', *Journal of Experimental Criminology*, 12.2, 187-207

Tompson, L., Bowers, K., Johnson, S.D., and Belur, J. (2015) EMMIE evidence appraisal coding tool and user guide. [Dataset] UCL Department of Security and Crime Science: London, UK.

Figures and tables:

Figure 1: The Crime Reduction Toolkit Landing Page

What Works
Crime Reduction **W**

BetterEvidence™
for Better Policing

ABOUT US RESEARCH CRIME REDUCTION TOOLKIT PARTNERSHIPS SUPPORT

Home / Crime Reduction Toolkit

Crime Reduction Toolkit

Intervention	Impact on crime	How it works	Where it works	How to do it	What it costs
	Effect	Mechanism	Moderator	Implementation	Economic cost
Alcohol ignition interlock	✓✓	⚙️	📍	?	£
Alcohol tax and price policies	✓✓	⚙️	📍	?	£
Alley gating	✓✓	⚙️	📍	?	£
Alternative education programmes	✗✓	⚙️	📍	?	£
CCTV	✓✓	⚙️	📍	?	£
Child social skills training	✓	⚙️	📍	?	£
Cognitive Behavioural Therapy (CBT)	✓✓	⚙️	📍	?	£
Cognitive Behavioural Therapy (CBT) for Domestic Violence	✓	⚙️	📍	?	£

Key

Quality of evidence

- No information
- Limited quality
- Moderate quality
- Strong quality
- Very strong quality

Filters

Impact on Crime
(select a range using the markers below)

Search term...

Violent Crime Off On

Property Crime Off On

Offender Based Off On

Victim Based Off On

Figure 2: Rating scales for the methodological quality of evidence in systematic reviews appraised using EMMIE

Effect

- 0 = Most forms of bias that could influence the study conclusions remain
- 1 = Although the review was systematic, many forms of bias that could influence the study conclusions remain
- 2 = Although the review was systematic, some forms of bias that could influence the study conclusions remain
- 3 = The review was sufficiently systematic that many forms of bias that could influence the study conclusions can be ruled out
- 4 = The review was sufficiently systematic that most forms of bias that could influence the study conclusions can be ruled out

Mechanism

- 0 = No reference to theory - simple black box
- 1 = General statement of assumed theory
- 2 = Detailed description of theory - drawn from prior work
- 3 = Full description of the theory of change and testable predictions generated from it
- 4 = Full description of the theory of change and robust analysis of whether this is operating as expected

Moderator

- 0 = No reference to relevant contextual conditions that may be necessary
- 1 = Ad hoc description of possible relevant contextual conditions
- 2 = Tests of the effects of contextual conditions defined post hoc using available variables
- 3 = Theoretically grounded description of relevant contextual conditions
- 4 = Collection and analysis of relevant data relating to theoretically grounded moderators and contexts

Implementation

- 0 = No account of implementation or implementation challenges
- 1 = Ad hoc comments on implementation or implementation challenges
- 2 = Concerted efforts to document implementation or implementation challenges
- 3 = Evidence-based account of levels of implementation or implementation challenges
- 4 = Complete evidence-based account of implementation or implementation challenges and specification of what would be necessary for replication elsewhere

Economics

- 0 = No mention of costs (and/or benefits)
- 1 = Only direct or explicit costs (and/or benefits) estimated
- 2 = Direct or explicit and indirect and implicit costs (and/or benefits) estimated
- 3 = Marginal or total or opportunity costs (and/or benefits) estimated
- 4 = Marginal or total or opportunity costs (and/or benefits) by bearer (or recipient) estimated

Figure 3: Hypothesised context-mechanism-outcome configurations for alley gates (from Sidebottom et al 2017a)

Context	Mechanism	Outcome
Offenders select targets with which they are familiar	Closing alleys removes vulnerable properties from likely offenders' awareness spaces	Reduction of domestic burglary, non-domestic burglary and theft from the garden and yards accessible from the alley
Alleys provide easy access to targets or a means of escape for offenders	Closing alleys increases the effort and risk of offending.	
Public alleys provide a legitimate excuse for would-be offenders to survey properties	Closing alleys removes excuses for loitering	
Open access alleys generate disorder and facilitate further crime and disorder	Closing alleys creates orderly space, providing cues to suggest that this is not a suitable place to offend because the risk is high	Reduction of drug use and dealing, prostitution, arson, the accumulation of litter, robbery and anti-social behaviour
Open access alleys are unregulated, no-one takes proprietary interest and disorderly behaviour goes unchallenged	Alley gates create defensible space where those backing onto the enclosed alleys display territoriality over it	
Alley gates are installed in high crime areas with little collective efficacy	The process of secure agreement for alley gates to be installed builds collective efficacy, enhancing mutual protection	
Areas known to have many rear alleys attract would-be offenders	Alley gating reduces the attractiveness of the neighbourhood for offenders who are looking for vulnerable targets	A diffusion of benefits

Table 1 - Descriptive statistics of the EMMIE-Q scores for 70 systematic reviews presented on the Crime Reduction Toolkit

	Min	Max	Median	Mean	S.d.
Effect	0	4	3	2.5	1.2
Mechanism	0	4	1	1.4	0.9
Moderator	0	4	2	1.6	1.0
Implementation	0	4	1	1.5	1.0
Economics	0	2	0	0.1	0.3