

What works in crime reduction Toolkit: Method statement

This brief note is intended to summarise the approach taken to create the What Works in Crime Reduction Toolkit. The principal aims of the toolkit are 1) to provide practitioners with a succinct summary of the conclusions of systematic reviews of crime reduction interventions; and, 2) to rate the quality of the evidence on which those conclusions are based, or the approach to analysis used to summarise it, so that users of the toolkit know which conclusions warrant more (or less) confidence. The method used to produce the toolkit was devised and implemented by researchers at the UCL Jill Dando Institute of Security and Crime Science and co-funded by the College of Policing and Economic and Social Research Council (ESRC grant ES/L007223/1). Development of the toolkit was undertaken as part of a programme of research involving a consortium of eight UK universities, in which ideas are freely exchanged. Many colleagues have thus contributed to this work and the College of Policing has been consulted throughout its development. The toolkit was developed in a number of stages as follows:

Stage 1: Searching for relevant systematic reviews

The first stage was to identify *systematic reviews*¹ of interventions that have been used to prevent crime. To ensure that the method used to locate the crime reduction evidence was systematic, transparent and replicable, a review protocol that outlined the procedures to be used to search for the literature was produced (Bowers et al, 2013). This protocol was reviewed by staff at the College of Policing and a panel of external experts. To qualify for inclusion a study needed to meet two specific criteria: 1) to be a systematic review and/or a meta-analysis and 2) to include a quantifiable outcome measure relating to crime. Hence, reviews that measured impact only on (intermediate) behaviours or non-crime outcomes, such as school attendance or self-reported level of aggression, were not included.

The search for systematic reviews involved 1) comprehensive searches of relevant databases using predefined search terms; 2) assessments of whether identified studies met pre-defined inclusion criteria; and, 3) regular inter-rater reliability exercises to monitor levels of agreement in determining which studies did and did not meet our inclusion criteria. This process yielded 325 unique systematic reviews that met the inclusion criteria. These

¹ A synthesis of the research evidence on a particular topic, which uses strict criteria to exclude studies that do not fit certain methodological requirements. Systematic reviews that provide a quantitative estimate of an effect size are called meta-analyses.

were then summarised and taken on to the next stage. Further details of stage 1 and a description of the resulting evidence base can be found in Bowers et al. (2014).

Stage 2: Devising a system to capture and assess the quality of evidence from systematic reviews

The intention was to produce a system that could be used to capture evidence that would be useful to practitioners and policymakers interested in reducing crime. Evidence was therefore sought on a range of topics relevant to crime reduction. It was deemed important that the evidence collected was not restricted to purely whether an effect was present, but extended to information that would assist future implementation decisions in a particular local context. Results from studies adopting both quantitative and qualitative approaches were used alongside one another.

A framework was developed to assess five dimensions of each systematic review, summarised by the acronym EMMIE. These refer to

- a) Effect size/s produced (the magnitude of any impact on crime),
- b) Mechanism/s activated (how interventions work),
- c) Moderators/contexts for the activation of the mechanism/s, (where it works best)
- d) Implementation conditions that supported/obstructed delivery (how to do it)
- e) Economic assessment of interventions (what it costs).

For each of the five EMMIE dimensions, two components were distinguished: EMMIE-E relates to the 'Evidence' that emerges from reviews (see Stage 4 below) and EMMIE-Q relates to the 'Quality' of that evidence. Both are needed for users of systematic review evidence to gauge what is or is not known and with what confidence.

Rating the quality of evidence in terms of its reliability and validity is fairly well established for estimates of effect size presented in systematic reviews. For example, good reviews increase validity by having a transparent and well-designed search strategy, featuring a valid statistical analysis, sufficiently assessing the risk of bias (e.g. do published and unpublished studies report different effects), considering the validity of the way outcomes are measured and/or combined, and conducting separate analyses to examine the influence of distinct evaluation research designs on estimated outcomes. They might also quantify an overall effect for unanticipated outcomes such as displacement caused by a geographically focused intervention. The extent to which systematic reviews address these issues was thus used to

rate them in terms of the quality of the evidence presented with regards to the estimated Effect size of an intervention.

Less well established is the rating of quality on dimensions other than effect size. It was therefore important to consider what elements would be necessary to demonstrate

sufficient care and attention to these. Taking Mechanism as an example, some reviews make no reference to how an intervention might work - simply assuming that it had been implemented and it would have some impact on crime. On the other end of the scale, reviews may be very detailed about the mechanism through which an intervention might bring about its effects. Such reviews might (for example) map out a casual chain of events that should take place to lead from an action (e.g. installing a burglar alarm) through to an impact on crime (e.g. a reduction in burglary). They might discuss different possible mechanisms and make statements as to the plausibility of these. Exceptionally, they might make predictions about what would happen if different mechanisms were at play and use empirical data to test, or partially test these. Many reviews fall somewhere between these two extremes. Hence, we use the rigor with which these concepts are explored to produce a system of calculating an EMMIE-Q score for Mechanism. Similar reasoning was used to produce Q scoring systems for the other elements of EMMIE. More detail for each of these can be found in Johnson et al (2015). However, for transparency we provide a narrative guide to the Q-scores (for evidence quality) for each dimension below.

EffectQSCORE

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

MechanismQSCORE

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

ModeratorQSCORE

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 variables that are at hand

3 = Theoretically grounded description of relevant contextual conditions

4 = Collection and analysis of relevant data relating to theoretically grounded moderators and contexts

ImplementationQSCORE

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

EconomicsQScore

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

Stage 3: Coding the studies

Crime clearly takes many forms; so too do methods intended to reduce it. Consequently, of the systematic reviews identified in stage 1, some focus on particular interventions (e.g. the impact of CCTV on crime, or the impact on offending of electronic monitoring), whilst others consider a series of interventions that may address a general problem (e.g. juvenile offending). The decision was made to concentrate on systematic reviews of the former kind initially since their singular focus provides more unequivocal evidence regarding a particular intervention. However, where further relevant evidence on the intervention in question was available in reviews of the latter kind, this was included in the material generated for the toolkit. In some cases, multiple reviews exist on the same intervention. Our process used the review of the *highest quality* as the basis for the evidence.

Where two or more reviews were of the same quality, we used the most recent. Where other reviews produced further reliable evidence, this was integrated. Where they provided no new reliable findings or were subsumed by others they were not included.

Two independent reviewers coded each review using the EMMIE framework. Coders compare their ratings and moderate their scores to agree a final set. Each dimension of EMMIE comprises a number of sub elements (see, Johnson et al., 2015), so a coding spreadsheet – which includes over 100 items – was developed to systematically capture each coder’s ratings (Tompson et al, 2015). To ensure the final quality rating generated from these 100 items was consistently rated across reviews, the spreadsheet has scoring rules (embedded as excel formulae) and automatically produces an EMMIE-Q profile for each review.

Stage 4: Translating EMMIE into a practical tool.

The coding spreadsheet captures how each reviewer scores each review but does not present this in a way that is easy to follow. The College of Policing and UCL therefore worked together to develop an online toolkit to present this information in a more accessible form. This resulted in two formats: a summary or ‘landing page’, and an associated ‘narrative’ that provides fuller details. The toolkit presents evidence at *intervention* level. When an intervention is covered by more than one systematic review the results are presented together in the toolkit. This includes relevant findings found in reviews of multiple interventions. Where information was obtained from multiple reviews, the evidence reported under each EMMIE dimension comes from the review(s) that attained the highest evidence quality score on that particular item, as this evidence is the most reliable. Where evidence from additional reviews is seen as important to the narrative, the additional review is clearly indicated, along with its quality score for that EMMIE dimension.

Some features of the tool include:

1. A landing page which presents the rating and ranking of interventions by impact, quality of evidence, implementation information and costs. For each intervention this page summarises the available evidence against each of the five dimensions of EMMIE. The landing page provides, for example, an indication of the measured impacts of an intervention (see below), the presence or absence of information on variations in effectiveness by context (the ‘moderator’ dimension) and the quality of

the evidence presented. In the case of the EMMIE dimensions, the tool indicates whether such information is available. For the Effect of intervention, a categorical scale is used to summarise whether the review(s) suggest that an intervention:

a) is associated with an overall reliable (in statistical terms) reduction (or increase) in crime across the studies reviewed; b) has been shown to reliably (in statistical terms) decrease or increase crime in some studies, or under some conditions, but not to do so consistently across studies; or c) never been shown to have a reliable impact upon crime.

The categorical Effect scale is presented in the form of ticks and crosses. A simple easy to read explanation of these ratings is provided [on the website](#) but the precise interpretation of these ratings, which was used to code the reviews, is shown in the table below. Many reviews employ statistical meta-analytic techniques to synthesise findings across studies. When conducted appropriately (see above), compared to simpler approaches (e.g. counting how many studies report a statistically significant effect), a statistical meta-analysis produces a more reliable estimate of the overall effect of an intervention by taking advantage of all of the available data. Thus, for studies that include a statistical meta-analysis, it is possible to indicate with some certainty if the intervention has had a reliable overall impact on crime across studies, and/or if it has done so in some studies or under certain conditions.

However, it is important to note that systematic reviews do not always include a statistical meta-analysis. For instance, many narrative reviews describe the results of individual evaluations but do not conduct a statistical analysis to compute an overall estimate of effect size. This can be for a variety of reasons, but in such cases any discussion of the overall effect of intervention may be unreliable. Hence, for such reviews, the coding applied is limited to an indication of whether statistically reliable effects have been observed in some studies or under some conditions. No indication is provided as to whether the intervention has had an overall effect on crime, as to do so would be misleading.

Number	Rating	Interpretation
		It is important to note that the evidence presented and its interpretation is based on the systematic review on which the effects of an intervention are based alone. Conclusions are not based on other sources of information.
1	X X v'v'	No evidence to suggest that the intervention has had a statistically significant impact on crime.
2	X X v'v'	Statistical meta-analysis suggests that overall, the intervention has had a positive and statistically reliable effect on crime.
3	X X v'v'	Overall, the intervention has not had a statistically significant effect on crime (or this was not tested), but there is evidence that it has had a statistically significant positive impact on crime in one or more studies, or under certain conditions.
4	X X v'v'	Overall, the intervention has not had a statistically significant effect on crime (or this was not tested), but there is evidence from one or more individual studies that it has had either a statistically significant positive or negative impact on crime, depending upon the conditions.
5	X X v'v'	Statistical meta-analysis suggests that overall, the intervention has had a statistically significant positive effect on crime, but it has also had a statistically significant negative effect on crime on one or more studies, or under certain conditions
6	X X v'v'	Statistical meta-analysis suggests that overall, the intervention has had a negative and statistically reliable effect on crime.
7	X X v'v'	Overall, the intervention has not had a statistically significant effect on crime (or this was not tested), but there is evidence that it has had a statistically significant negative impact on crime in one or more studies, or under certain conditions.
8	X X v'v'	Statistical meta-analysis suggests that overall, the intervention has had a statistically significant negative effect on crime, but it has also had a statistically significant positive effect on crime in one or more studies, or under certain conditions.

2. A narrative for each intervention, summarising the evidence on each dimension of EMMIE. These narratives are constructed by at least two researchers and edited by staff from the College of Policing. The findings reported are limited to those included in the particular review in question. No additional searches were conducted to identify subsequent evidence that might challenge the findings of the review, nor was opinion sought about the reviews. Exceptionally, where additional important evidence is (or was made) known to the review team, this is usually included in a 'general considerations' section and the source is clearly marked.
3. A filter is included (on the landing page), enabling users to find relevant interventions by type of crime outcome (e.g. violent crime), type of population targeted (e.g. locations, victims or offenders) or strength of evidence on effect size.

Dated: February 2015

References

Bowers, K., Johnson, S., Tilley, N., Tompson, L. and Belur, J. (2013). Protocol for Work Package 1: Identifying Existing Systematic Reviews of Crime Reduction. [Available here.](#)

Bowers, K., Tompson, L. and Johnson, S. D. (2014) Implementing Information Science in Policing: Mapping the Evidence Base Policing; doi: 10.1093/police/pau058. [Further details available here.](#)

Johnson, S.D., Tilley, N., and Bowers, K.J. (2015). Introducing EMMIE: An evidence rating scale to encourage mixed-method crime prevention synthesis Reviews. [Structured abstract](#); Journal of Experimental Criminology, forthcoming.

Tilley, N. (2015) EMMIE as a framework for inclusive evidence appraisal. In preparation.

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.