

Street-level Theories of Change: Adapting the Medical Model of Evidence-based Practice for Policing

Nick Cowen, New York University School of Law and Nancy
Cartwright, Durham University and UCSD

CHES Working Paper No. 2019-02

[Produced as part of the Knowledge for Use (K4U) Research Project]

Durham University

February 2019



The K4U project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 667526 K4U) The above content reflects only the author's view and that the ERC is not responsible for any use that may be made of the information it contains

Street-level Theories of Change: Adapting the Medical Model of Evidence-based Practice for Policing

Nick Cowen and Nancy Cartwright

Nick Cowen
New York University School of Law
40 Washington Sq. South
New York
NY 10012
nick.cowen <at> nyu.edu

Nancy Cartwright
Philosophy Department
Durham University
50 Old Elvet
Durham
DH1 3HN
nancy.cartwright <at> durham.ac.uk

Abstract: Evidence-based medicine (EBM), with its evidence hierarchies and emphasis on RCTs, meta-analyses and systematic reviews, sets the model for evidence-based policy almost everywhere, policing no exception. But how closely should policing follow this model? We argue that RCTs can tell you little about what you need to know for real-world practice: will this policy work where and when you implement it? Defending that it will do so takes good theory. For RCTs to play a role in theory development, they must be set into a larger body of knowledge, including local knowledge about the sites of implementation. Unfortunately the standard EBM model generally ignores the other kinds of knowledge needed. An alternative model for evidence-based policing, similar to that of the new movement for 'EBM+' and immanent in the practice of realist synthesis, focuses on the arguments that proposed policing policies will work where and when they are implemented and looks for the evidence needed to support those arguments.

Evidence-based policing policy (EBPP) has become a prominent feature of criminal justice research and increasingly influential in policing practice. Is its influence in its current form justified? Proponents of EBPP draw on the apparent success and consensus surrounding scientific medicine and the supposed link between this success and an emphasis on experimental methods. In the United Kingdom, the evidence-based paradigm gained greater prominence when the Conservative-led Coalition Government offered a platform of austerity with a promise to 'do more with less' in public services, with cuts focused on the criminal justice system (Bandyopadhyay, 2013). As a result, several critics see it as a way of imposing ideologically inspired policies under the cloak of scientific neutrality (Lumsden and Goode, 2018).

The argument of this chapter is as follows. Evidence-based medicine was an attempt to reform a more traditional medical model of clinical expertise in judging the efficacy of medical interventions. Yet the practical success in doing so is not chiefly the result of EBM narrowly construed but rather medicine's relationship with much broader scientific theory. We will focus on one specific kind of theory: what in medicine are called the 'mechanisms' by which treatments work. A mechanism is a kind of medical causal model that shows step-by-step the theory of how the treatment is supposed to produce the targeted outcome. In evidence-based practice more generally, these kinds of causal models often come under the label 'theories of change' (TOC). In policing, as opposed to medicine, a practical TOC will inevitably contain quite a bit of informal theorizing: approximate, defeasible generalizations of human behaviour that are nevertheless useful for figuring what is likely to work to reduce crime in context. We call these 'street-level theories of change' (STOC). Experimental research can certainly help support the use of such theories, alongside other forms of research and sources of knowledge, but cannot justify the generalized use of supposedly 'evidence-based' interventions.

The importance of mechanisms in EBM has long been disputed, with 'mechanistic evidence' often appearing near the bottom of a hierarchy of evidence types that warrant conclusions about the efficacy of an intervention. But this is now changing in EBM. Partly in response to the kinds of weaknesses in more conventional evidence-based practices, proponents of 'EBM+' have been taking a step back from a narrow experimentalist paradigm and recast EBM as a more general endeavour using key evidence types. They argue strongly for supplementing data analysis with an examination of the mechanisms through which interventions are supposed to work (Marchionni and Reijula, 2018; Parkkinen et al., 2018). Our recommendation to focus more on STOCs for

policing evidence is exactly in line with these ideas from EBM+. STOCs, we urge, should be a key ingredient in EBPP+.

Understanding the relationship between EBM and EBP is important because the push for experimental designs is strongest in medicine, yet clinical proponents of EBM have seen their approach as applicable beyond medicine from the beginning (Cochrane, 1972). Clinicians today are influential backers of remaking policy in other sectors in this image (Goldacre, 2015; Haynes et al., 2012) and the institutions most closely associated with EBM, for example the National Institute for Health and Care Excellence in the United Kingdom, are frequently held up as a model for other sectors to adopt. Yet EBM as a policy initiative is itself a contextual response to social conditions and professional preoccupations. These must be acknowledged if we are to grapple with the strengths and weaknesses of EBPP in its own context.

So we begin with a brief outline of the history of the development of EBM as part of the critique of a previous medical model and some of the intellectual connections between it and EBPP. This is the most intuitive way we have found for outlining the difficulties with applying a medical model to policing, reflecting the context in which EBM was originally developed and then disseminated. It helps us to distinguish three key kinds of relevant limits:

1. There are contributors to good health that critics of the medical model highlight that EBM does not attempt to address. There are equivalent contributors to a humane and lawful society that EBPP cannot address.
2. There are features of classical EBM, especially a lack of attention to mechanisms, which limit the use of such research in virtually any domain. These features are addressed under the recently developed paradigm EBM+. It would be useful for policing research to move in the direction of an EBPP+ and we think theories of change can help better highlight the necessary understanding of mechanisms for figuring out what works in context.
3. There are aspects of crime that can only be examined with middle-range theories, not explicit scientific theories as for much of EBM+, and this changes what we can expect a theory of change in policing to look like. So we suggest the label STOCs for exploring typical mechanisms in policing.

The Medical Model and its Discontents

During the 1970s medicine was widely considered to be facing an intellectual crisis (Alvarez-Dardet and Ruiz, 1993). Part of this was a generalized suspicion of traditional social institutions.

Part of it was due to scholars critically examining new evidence about the effectiveness of clinical practices.

R.D. Laing (1971: 38–39) popularized the term ‘medical model’ to critique what he saw as unwarranted confidence in clinical expertise, especially with regards to mental health. Laing’s ‘medical model’ is a description of what he saw as the professional and social practices of doctors when interacting with and managing patients. It includes familiar stages in medical practice, such as hearing initial complaints about illness, establishing case history, physically examining patients, ordering specific tests, diagnosis, prognosis and treatment. Laing used this concept of the medical model to argue that doctors should not use their generalized authority as clinical practitioners to engage in practices that are not, in fact, scientifically warranted (Engel, 1977). A few years later, the social critic Ivan Illich (1982) wrote a broader popular critique of scientific medicine as cultural iatrogenesis: the replacement of effective traditional communal practices of dealing with ailments with harmful, exploitative medicine.

Another critique sought explanation for variations in health not from medical practice but from changes in social and economic factors. Thomas McKeown (1976) argued that the main cause of improved health, especially a dramatic drop in mortality in the modern era, was economic growth leading to better nutritional standards and improved social conditions. This thesis calls into question the very basis for associating the growth of scientific medicine with better human health. Contemporary research suggests that economic growth was indeed comparatively important in the 19th century as it increased access to nutrition, while public health (especially access to clean water, waste removal and improved personal hygiene practices propagated by public institutions) were more significant in the 20th century (Cutler et al., 2006; Deaton, 2015).

Both these approaches have analogical perspectives in crime reduction. A patient-centred, holistic approach to medicine that Laing affirms parallels approaches to crime prevention that attempt to address the needs of the whole person rather than just offending behaviour, as well as to resolve problematic relationships and community failures that are associated with underlying causes of crime. An equivalent perspective to McKeown’s explains predominant trends in crime not by police and criminal justice responses but with socio-economic factors such as wealth, poverty, inequality, unemployment (Han et al., 2013; Witt et al., 1999), as well as broader social determinants such as education, health and geographic mobility.

We believe that both these approaches are valuable, and it is unfortunate when policymakers ignore them. Evidence-based policing, however, focuses on individual and situational

interventions that field professionals might plausibly be able to implement. Without denying the importance of more systematic approaches to crime reduction, here we focus on exploring weaknesses that are internal to the evidence-based paradigm (and how to ameliorate them).

Evidence-based medicine as a radicalization of the medical model

Just a year after Laing invented and attempted to slay the 'medical model', Archie Cochrane (a colleague of McKeown's) published *Effectiveness and efficiency: Random reflections on health services* (1972) a seminal critique of existing clinical practice from a different perspective. His focus was on a puzzle he observed while working at the Nuffield Trust (then the Nuffield Provincial Hospitals Trust). As the National Health Service grew, healthcare outputs did not appear to be improving in line with increased resources (Cochrane, 1972: 13).

Cochrane's answer was that this observed inefficiency in the NHS was due to clinicians' hostility to using experimentally proven efficacious techniques and unwillingness to conduct randomized trials as part of their own practice to extend public knowledge of scientific medicine. Cochrane (1972: 20) was particularly critical of the use of expert medical opinion that he felt was unjustifiably affirmed in public health debates. He regarded observational study designs as unsatisfactory, believing they often produced results that misled medical practitioners. He further warned that doctors were often 'easily bemused by evidence from what appears to be a more basic science' (1972: 30), that is treatments based on theories tested in some scientific domain but not in clinical practice itself. He was annoyed that the notion of 'experiment' was applied promiscuously, amounting in common discussion to any observation of an innovative practice regardless of how the results of the experiment were supposed to be inferred.

For Cochrane, only randomized-controlled trials offer satisfactory evidence of causal efficacy of a medical intervention because they allocate treatment independent of human judgement (1972: 22) and in the case of double-blind trials, conduct and observe treatment independent of clinical knowledge and judgment. So advantageous in Cochrane's view is this research design for discovering what works in medicine that he sought explanation for why they were not more widely used. He concluded that one key problem was that conducting RCTs was fascinating and productive for principal investigators but that the meticulous protocol-following and record-keeping was boring for the clinicians involved.

Cochrane, and his work, inspired a generation of clinical reformers (Chalmers, 2008). Responding to Cochrane's concerns (1972: 63) that obstetric interventions were particularly lacking in evidence, Iain Chalmers helped to pioneer an early electronic database of experimental studies in

that area (Starr et al., 2009). The database and associated institute grew in scope to become the *Cochrane Collaboration* (now simply called *Cochrane*), one of the world's most well-known institutes for the promotion of EBM and information resources for the conduct and collation of systematic reviews of medical interventions. It inspired the *Campbell Collaboration*, an equivalent review archive and institution in social policy that now similarly enjoys international recognition. This body of research is also a significant contributor to the work of the National Institute for Health and Care Excellence (NICE) that sets guidelines based on efficacy and cost-effectiveness for the National Health Service.

Seen from a historical perspective, EBM is one reaction to a crisis in confidence in medicine that also prompted the radical critiques of Laing and McKeown. Like Laing, proponents of EBM are concerned about unjustified reliance on clinical authority rather than solid evidence. They do not want medical decisions taken based on what they see as capricious opinion elevated to institutionalized power. Unlike Laing, EBM proponents do not wish to abolish or curtail the medical model as such, but rather to refine, perhaps even to 'radicalize' or 'purify' it. This medical model is not one of experienced professionals using a mix of formal knowledge, shared practices among peers, and personal intuition, to develop a treatment path for individual patients. Instead, on the EBM account, medical judgement should be guided predominantly by the results of properly conducted medical research defined as offering a strong independent basis for causal inference between treatment and outcome (Guyatt, 1992). Only after this rigorous evidence is taken into account do EBM proponents acknowledge a role for individual clinical judgement (Sackett et al., 1996).

Laing's focus was on the relationship between the doctor and the patient. Cochrane's (and the EBM movement's) approach is prompted, in addition, by the relationship between medical practitioners and the government (and the wider public interest). Medical practitioners are stewards of limited public resources. For EBM proponents, it is important that patients are not misled by their doctor's unwarranted intuitions about how best to treat them. It is equally important that public resources are not squandered. As a result, contemporary EBM proponents seek to identify not just efficacious treatments, but cost-effective treatments. We can sum up this contemporary evidence-based medical model as having three characteristics:

1. Conducting studies with a focus on establishing a strong, direct causal basis for treatment and improved outcomes, with randomized-controlled trials presented as the 'gold standard' of such studies.

2. The use of systematic reviews with pre-registered search protocols and meta-analysis aimed at properly considering all the evidence together to produce a credible summary result, with a focus on attempts to identify and correct for bias in publication.
3. The use of economic models and cost-benefit analyses to recommend standard treatments to NHS hospital trusts in the United Kingdom and guide public and private insurers in the United States.

Note especially that theory does not enter into consideration in EBM. Yet medical treatments are generally highly theorised. Consider pharmaceuticals. It is hugely expensive to conduct RCTs so these are performed only on drugs that are judged highly promising. Part of that promise may be based on experiments on model organisms. But before that pharmaceutical companies will have engaged in a great deal of biotechnology and pharmacology using detailed biochemical and other theories to identify compounds that might have the right characteristics. Thereafter they will have continued a heavy investment in research and theory application to develop the drug, paying serious attention to the detailed theories that support that it can do what they hope, and how it can do it.

In general, that a claim is backed by good theory is taken to as a strong reason in favour of it. But EBM does not allow this to count as evidence that a treatment will work. It may help us *discover* a treatment but it appears nowhere in EBM ranking schemes for the kind of thing that can *confirm* that the treatment works. EBM+, as we discuss below, takes a broader perspective. All efficacy claims are dicey and all studies will have some flaw or other and almost all will have been conducted on populations different from those who will be treated with the drug. Given that, we should look at as much evidence as possible, and especially at evidence of different types. So both mechanisms and trial evidence matter. We urge this is true in policing as well. But laying out the mechanism – providing a theory of change for the intervention – does much more than help with warrant, as we shall explain.

Professional policing and social science

The development of modern medicine and modern policing share some parallels. The professionalization of both medicine and criminal justice is correlated with significant improvements in both public health and security: reductions in mortality and reductions in homicide (Deaton, 2003; Eisner, 2003). However, as in medicine, it has proven a challenge to determine with any certainty the causal contribution of professional policing to these changes in

crime in society. But the experience gave policymakers confidence in the principle of rational reform. This led to many re-workings of policing over the course of the 20th Century enabled by technological changes and inspired by a variety of theories of social change.

Starting in the 1970s, in response to some pessimistic evaluations of crime reduction interventions and increased crime in some urban areas, the notion that 'nothing works' took hold of criminology: attempts to change the life courses of offenders were taken to be helpless in the face of broader structural and sociological factors (Martinson, 1974; Cullen and Gendreau, 2001; Weisburd et al., 2017). Although this pessimism was most concentrated around the issue of rehabilitation of prisoners, the intuition spread to the capacity of the police to address crime. The truth of this contention is at least contestable. In addition, policymakers and the police found this perspective to be unhelpful, especially at a time when crime reduction and safety was forming an increasingly important part of political platforms. There was a gap between government priorities and criminological research. It is possible that this gap is what allowed plausible, popular but ultimately poorly formulated aggressive policing interventions, including some forms of 'broken windows policing' and scared-straight programs, to become the default approach for politicians by the end of the century (Braga et al., 2015; Petrosino et al., 2003; Weisburd et al., 2015).

Evidence-based policing as a solution to criminological pessimism

The stage was set for a revived, pragmatic, stakeholder-oriented approach to policing research. This was a role that an EBM-inspired research program has attempted to fill. The new wave began in the United States, but with involvement from British academics. Lawrence Sherman and colleagues (1997) at the University of Maryland produced a report for the United States Congress on preventing crime. It was an early and important exponent of this approach. They drew on Cook and Campbell (1979) for their methodological inspiration, who, in turn, had drawn substantively from Cochrane. Sherman *et al.*'s results were tentative. Their strongest conclusions were that crime reduction initiatives should be focused on high-crime urban areas, and (now a common refrain) that a great deal more research evaluating policing interventions was needed.

A special issue of *The Annals of the American Academy of Political and Social Science* brought together many of the themes and contributors to the developing evidence-based movement in criminal justice studies both in the United States and United Kingdom. Welsh and Farrington (2001) argued that randomized studies provide the most convincing evidence in principle, before acknowledging that the nature of criminal justice means that a wider range of study designs must be accepted. Farrington and Petrosino (2001) introduced the official *Campbell Collaboration Crime*

and Justice Group, explaining they would support and collate systematic reviews of studies based on explicit objectives, eligibility criteria, pre-registered unbiased searches. The aim was to produce quantitative measures whenever possible and appropriate. Wilson (2001) described meta-analytic approaches as offering a way of summarizing the results of the academic literature in way that was understandable to policymakers.

Subsequently, Farrington *at al.* (2002) refined their approach into the now widely used *Maryland Scientific Methods Scale* (SMS), a tool for categorizing the alleged strength of a study when conducting systematic reviews. Level 5 studies are randomized experiments. Level 1 studies offer only cross-sectional statistical correlations between comparable areas, with higher levels more closely imitating the randomized designs at the top. More recently, Sherman and colleagues (2016) have developed the cost-benefit analysis feature of EBP with the introduction of a Crime Harm Index that attempts to show where policing resources can best be applied to improve social welfare in the most efficient ways possible.

The limits of the evidence-based paradigm for medicine and policing

Is the special emphasis on RCTs as an ideal against which to measure other research designs for causal claims justified in the way that Cochrane and contemporary EBP proponents claim? Our view is that RCTs can be useful for extending clinical and professional knowledge, but they have no claim over other rigorous approaches to causal research (Cartwright, 2011). Randomization does not represent an ideal for establishing causal efficacy as implied by its frequently imputed 'gold standard' status among research designs (Bothwell et al., 2016; DeMets and Cook, 2018; Jones and Podolsky, 2015; cf. Cartwright, 2007a).

Why exactly would you want a research design to be based on randomization? You want to randomize when you do not know all the other factors that determine the outcome of an intervention and have no other way of controlling for them. RCTs are an attempt to accommodate the fact that you cannot conduct a relevant laboratory-controlled study in a field environment.

Basic scientific research involves controlled laboratory experiments that allow scientists to decide precisely what goes into the experiment. The result, in principle, is a recipe and set of instructions to produce a predictable outcome that indicates the way a causal mechanism works (for example, the role of a gene in expressing eye colour). These sets of instructions, shared among peers in academic publications, ought to be replicable by other laboratories where people are sufficiently

familiar with typical experimental techniques. When the instructions are properly followed (sometimes after a bit of trial and error when getting to grips with a new technique), they ought to produce the same results with precision every time. In such cases, scientists do not worry about sample size to confirm whether the results are credible or reproducible.

No such precision, of course, is possible or expected in the treatment of human beings in the field. Human physiology, individual behaviour, and the contexts in which people live vary in a great deal of ways that could potentially alter the outcome of an intervention from person to person and context to context. Ultimately the dimensions of variation are limitless. An RCT is a way of testing something for which an environmentally controlled study is infeasible or impossible. It cannot credibly claim to offer a complete set of instructions that will deliver the same result in each individual case. Instead, it looks at cases that differ in other causal factors than the one studied. What we can get from an RCT is an estimate¹ of the *average* treatment effect in the population studied. What value of the outcome an individual has depends on what other causal factors they are subject too. The RCT result averages across these. So we can expect the same average effect only in populations where these others are distributed similarly, or more precisely where they have a distribution that yields the same average.

Proponents of RCTs suggest that this average effect becomes more likely to predict, at least, the directional impact of the intervention (positive, negative or neutral) across different settings and populations the larger number of tests and the more widely the intervention is tested across contexts. Each new study builds and refines the evidence base. This can help explain the emphasis on regularly updated systematic reviews of existing experimental studies. Although not openly stated as such, the idea seems to be that good interventions will have an underlying central tendency for the outcome of interest. That central tendency, equivalent to something like the gravitational pull of a planet on a moon, is hidden by perturbations (like asteroids passing through the solar system) that disrupt any single observation. Similarly, no single study on a single study population can offer a strong claim about a tendency. But supposedly if you make enough observations, the average of multiple studies will supposedly point to the 'real' underlying tendency of the intervention. Even if the size of the effect – which depends on the distribution of the other causal factors in a population – differs from population to population, there seems to be the assumption that repeated positive effects sizes across varied population allows us to conclude that the treatment is on average causally positive.

¹ An *unbiased* estimate.

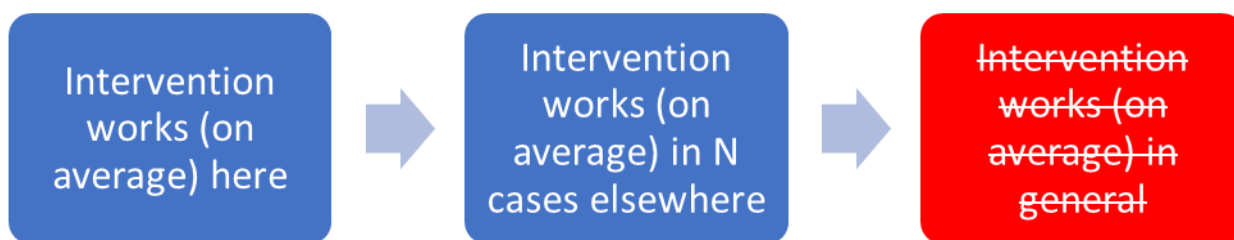


Figure 1: Inductive Reasoning

Unfortunately, such an approach falls foul of the perennial problem of reasoning via induction by simple enumeration. The population of interest where any intervention might be used is open-ended and usually undefined. The populations available for sampling, by contrast, are typically limited by all kinds of contingent restrictive conditions about who can be included. This is both for ethical reasons – we should not experiment on children, the elderly, the vulnerable or those who cannot give informed consent, and in aid of identifying the effect of just the treatment, which could be obscured by interaction with co-morbidities, other drugs, diet, environmental factors and the like. But it is limiting when it comes to estimating whether the treatment will work on other populations in different places with different characteristics. Cochrane himself complained about the lack of interest in randomized studies for populations in Catholic and Communist countries. The measured effect sizes in different populations can differ, often dramatically, from each other and then again from some average yielded by a meta-analysis of the averages across different test populations.

Even if these practical issues were surmountable, existing studies cannot sample future populations living under new social and environmental conditions. Sample populations inevitably deviate systematically from the population that clinicians are interested in for their practice. In the classic ‘black swan’ scenario, the appearance of black swans was surprising because they are native only to Australia, a landmass previously inaccessible to European naturalists. Before then, it was presumed from all previous observations that all swans were white.

Because it is recognized in medicine that the target populations who will be treated differ, there is much stress in medicine on *pragmatic trials* – ones done on ordinary population’s treated in ordinary circumstances in the ordinary way by ordinary clinicians. But the same problems loom: studies can show results only about the populations studied, and population characteristics vary. One may think: if we get a positive effect size at least we have learned that using a less skilled

clinician doesn't block effectiveness. But that does not follow. There may well be things in this new environment that offset the negative effect of the less skilled clinician.

Do these systematic deviations matter practically? They may or may not. It depends on whether the inevitable deviation is associated with a factor that impacts the intervention. What it means is that you cannot rely on any generalized average effect size to hold widely or in any specific new setting on the basis of experimental studies alone. The deviation between observable and unobservable cases is inevitably not random.

Reliance on attempts to generalize through empirical testing has two major problems. First, it produces over-confidence that an intervention that works 'on average' is worth doing in the case at hand on the presumption that there is some consistent underlying positive effect even if unobserved. That effect might be negated in the circumstances that the intervention is being implemented. Second, it can encourage practitioners to reject interventions that could really work this time if the setting is correct. Suppose that an intervention fails to work or produces a negative outcome in the average of the cases where it has been observed. From an EBP standpoint, that seems to make the intervention unpromising and, before long, not even worth further testing. But it could be that an intervention that does not work in general might be precisely the most effective in a new case. You can only know this if you have an understanding, or a substantiated theory, about why something works.

RCTs can contribute to the tangle of theory and evidence that establishes the capacity for an intervention to achieve an intended outcome, or at least to set a process in motion. In other cases, a combination of well-established theory and simple observation can be enough to establish capacity without any RCTs. Smith (2003) illustrates this point by suggesting facetiously that there is no 'causal evidence' of the effectiveness of parachutes to prevent human fatalities when falling from thousands of feet.

The upshot of this is that EBM improves medicine and clinical practice through engagement with a changing framework of much broader scientific knowledge. It is one of many forms of research that develop our understanding of causal capacities and of how interventions are meant to produce their targeted outcomes. Interventions cannot rely on the warrant of past success in specific circumstances alone. They must be based on theories that explain when treatments are appropriate and when they are not.

Our critique applies to the traditional theory of EBM which underplays the importance of supporting theories in real medical science and practice. The same critique applies to all other experimentalist formulations of evidence-based practice, including policing. Policing involves interventions in a complex, adaptive, social field, rather than in predictable or controlled conditions. As in medicine, the experimental paradigm in policing aims to establish ‘what works’ through repeated testing of interventions with the expectation that testing in enough areas and statistically synthesizing the results will produce a general conclusion: i.e. a robust prediction about what will work in areas that have not been tested. The key problem is that there are limitless factors that will condition whether an intervention will work in context. What kind of factors should practitioners be aware of and which can be ignored? The answer to that will depend on the causal model underlying the premise of the intervention. Where does that account come from? Only from broader theoretical understanding drawn from criminology, psychology, sociology, and economics.

The key idea is this. When relatively small changes in context or implementation modify the effect of an intervention in practice, experimental studies may end up providing a spurious assurance of certainty in one dimension: the causal contribution of one example of an intervention, at one place, at one time with one set of people. In such cases, an alternative study that measures or accounts for the variety of factors at work can be far more useful, even if the information on which it is based is less secure. RCTs, by definition, attempt to ignore those factors rather than model them. Other experimental research designs that imitate RCTs can similarly lack a focus on what’s really needed for practitioners in policing.

A Street-level Theory of Change approach to evidence-based policing

We have argued that the value of EBM lies not in the capacity of experiments to estimate a general effect of an intervention but as one important way, among others, of substantiating theories about how to better treat disease. The question is, can evidence-based practice make a similarly compelling contribution to policing? What is a realistic goal of a rigorous approach to testing an intervention that parallels the sort of reforms emerging within the EBM movement?

Following what Cartwright and colleagues have proposed for social interventions more generally (Cartwright, 2007b, 2013; Cartwright and Cowen, 2014; Cartwright and Hardie, 2012; Cowen et al., 2017; Deaton and Cartwright, 2018; Joyce and Cartwright, 2018), we propose that evidence can still contribute to successful policing interventions in a way that better reflects the boundaries of

possible knowledge. It is using evidence to develop a defeasible causal model that will tell us whether and why an intervention will work. Rather than the inductivist approaches that implicitly underlie the experimentalist paradigm, our approach is based on the premise, following Mackie (1965, 1980) in philosophy and Rothman (1976) in epidemiology, that causal claims are INUS conditions: insufficient but necessary parts of an unnecessary but sufficient condition (INUS) to produce an effect. That they are unnecessary means that there are normally several independent paths through which an effect could happen, so no one of them is the only way to get a contribution to the effect. However, for any given condition that is expected to produce an effect, there will be further individual elements – supporting factors – that are necessary for that cause to produce the effect. Each of those factors that act together then are necessary parts of a whole complex that is enough (sufficient) for a contribution to the effect.

To predict if an intervention will work in some targeted new setting, it is a good idea to build a causal model for that setting. This is easiest done when programme developers have already provided a general TOC for their intervention. This is a theoretical model that explains, step-by-step with some precision, how an intervention is supposed to work to produce the desired outcome.

This TOC is important because it helps prompt a focus on two often-overlooked issues. First is the activity, and the underlying principle or mechanism, that is supposed to move the intervention from one step to another. Once we see what must happen at each step for the intervention to produce the final outcome, the second issue that emerges is the necessary conditions that must be in place at each step for that step to produce the next. Thus we begin to learn what the features are of an individual or context that matter to the chance of the causal process initiated by the intervention being able to carry through from start to finish.

At each step of a TOC there are local support factors that must be in place for the activity to work as intended. Their absence is something that must be fixed in addition to implementing the intervention, or ultimately the intervention must be abandoned for something that does not rely on such support factors. At the same time, the presence of local derailers (off-setters and interrupters) can diminish, negate or reverse the effect of the activity. The presence of support factors and the absence of derailers make up the causal model that allows practitioners to get a good idea whether and why an intervention will or will not work in a case.

At best only some of these will appear in a programme's general TOC, in part because what concrete details will help or derail the process at each stage will differ so much from context to

context. Of course, sometimes general advice is available – striking a match produces a flame only if there is *oxygen* present. But often in social settings such general descriptions will be so abstract as to be practically useless, sometimes amounting to little more than ‘Do all else necessary’. That’s why the program’s general TOC is usually only a starting point for building a more informative local one.

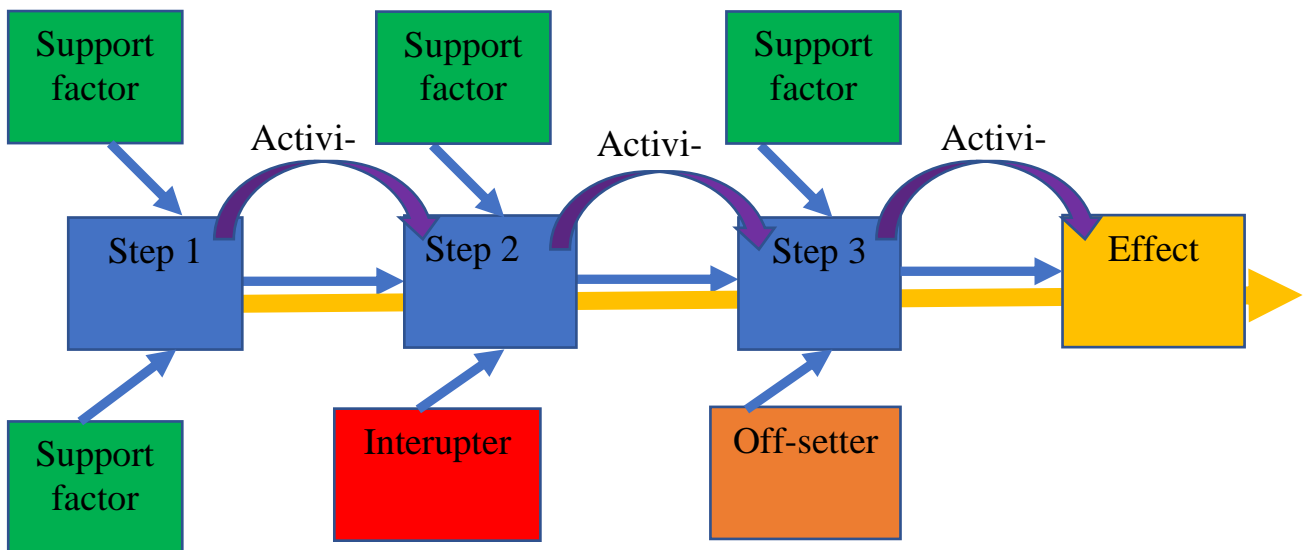


Figure 2: Street-level Theory of Change (STOC)

Figure 2 illustrates what is new in our approach to a TOC. The long arrow going all the way from the first step to the effect is equivalent to the direct intervention/outcome measure that a purist experimentalist approach implies with no concern about mechanisms. The steps in the square boxes look like a more familiar TOC. What we add is a focus on activities (denoted by curved arrows) that explain more precisely how you expect to get from one step to another. These activities refer to processes that one can see repeated in many contexts, not just in specific interventions. We also add effect modifiers (both support factors and derailleurs) that constitute the context in which the intervention unfolds.

The key difference between policing and medicine is the kind of theories used to justify our confidence in the activities at each step. Medical science draws from a wealth of scientific resources, including anatomy, molecular biology, bio-chemistry, physiology and epidemiology, and it uses a range of research designs to hypothesize and test models of disease and reactions to treatment. By contrast, most of the interventions that policing uses are based on theories of human behaviour. Human action and motivation is too complex to break down into precise,

predictable elements as one can more often do for many processes in the natural sciences. In addition, humans act strategically, reacting and adapting to changing novel circumstances. So instead of something like natural laws, these middle-range theories are more in the form of maxims (Pawson, 2000). In policing, we might call these 'street-level' theories. So what we are generally offering with these accounts is a Street-level Theory of Change (STOC).

Such street-level theories include: offenders prefer to commit easy crimes rather than hard crimes (Bowers et al., 2005); offenders usually wish to avoid apprehension (Abramovaite et al., 2018); offenders have habits and repeat strategies (Johnson et al., 2009); offenders are influenced by peers and family relationships (Abramovaite et al., 2015; van Mastrigt and Farrington, 2009); consuming alcohol or drugs lowers inhibition (Boles and Miotto, 2003). These theories are subject to inconsistencies and exceptions (Cartwright, 1983: 51–52; Elster, 1998). For example, adherence to common norms in a peer group may explain offending behaviour. But ignorance of or explicit deviance from norms may explain crime in other cases. Nevertheless, these theories apply in enough cases to supply consistent and reliable activities to make up a theory of change.

Positive deviance from the medical model

The weaknesses of a narrowly experimental evidence-based paradigm and the relative strength of a theory of change paradigm can be illustrated by the first systematic review commissioned by the *College of Policing*. It is on the installation of alley gates to reduce crime, especially burglaries (Sidebottom et al., 2017). According to EBPP theory, this study commits several errors in failing to establish causal relations according to a supposedly rigorous research design. In terms of practical usefulness, by contrast, the study has great value. On our account, it is precisely in the ways the study departs from the traditional EBM paradigm to that of EBM+ and further onwards to our idea of STOCs that it produces practical guidance. The parts that attempt to follow EBPP fail to measure up to it while not really contributing that much in terms of establishing the success of an intervention.

The premise of this intervention is that alleyways are an ambiguous feature of many neighbourhoods. They are neither public streets where suspicious behaviour is easily observed, nor are they purely private property where trespasses are easily identified. Instead, they are normally communal sites where an overlapping and changing group of people have legitimate reasons for being there. This facilitates burglaries by, for example, providing a more secluded exit for people carrying bulky valuable goods out of a private household.

Installing alley gates helps to clarify where unknown members of the public are meant to be found. If the gates are kept locked consistently, then the residents can physically restrict access to individuals they have invited. Even if they are sometimes unlocked, the gates represent a symbolic threshold beyond which members of the public are not supposed to cross without invitation from a resident. That means residents might be more willing to ask strangers to identify themselves and to notify police of suspicious activity.

The study combines a meta-analysis of studies with a realist-inspired approach based on an expanded review framework that supplements the Maryland SMS: EMMIE (a search for the Effect, Mechanism/Mediators, Moderators, Implementing features and Economic costs of an intervention) (Johnson et al., 2015). Although not the same as the STOC approach we have set out, EMMIE draws inspiration from several critiques of traditional evidence-based policy, including Cartwright and Hardie (2012), and so addresses some key concerns that we identify. Both our approach and realist syntheses turn appropriate attention to the activity² through which an intervention is supposed to work (rather than just that something supposedly works empirically), as well as the context (moderators or support factors) that have to be in place for something to work (Pawson, 2006; Pawson and Tilly, 1997). Figure 3 is a stylized illustration of what a STOC could constitute that covers several (though not all) the factors covered by Sidebottom et al. (2017).

Alley gating is an area-based policing intervention rather than an individual-level intervention. As a result, there are no RCTs available for this intervention. Initially following the Maryland SMS, Sidebottom et al. (2017) look to before-and-after study designs. These produce impressive results in terms of crime reduction. However, many take place at roughly the same time during a period when burglary reduction was a public policy priority. It is possible that some of these results are catching a secular reduction in property crime. Then the authors consider some quasi-experimental studies that compare areas non-randomly selected with matched control areas. Some of the studies also look at broader catchment areas that contain the areas provided with alley gates to test whether the intervention displaces crime into other places. In line with other studies (Weisburd et al., 2006), they find little evidence of displacement, suggesting that the alley gates produced a real net reduction in crime.

² Realists call this a 'mechanism'. They are thus using 'mechanism' differently than the step-by-step meaning of it we have been using and that is used by EBM+.

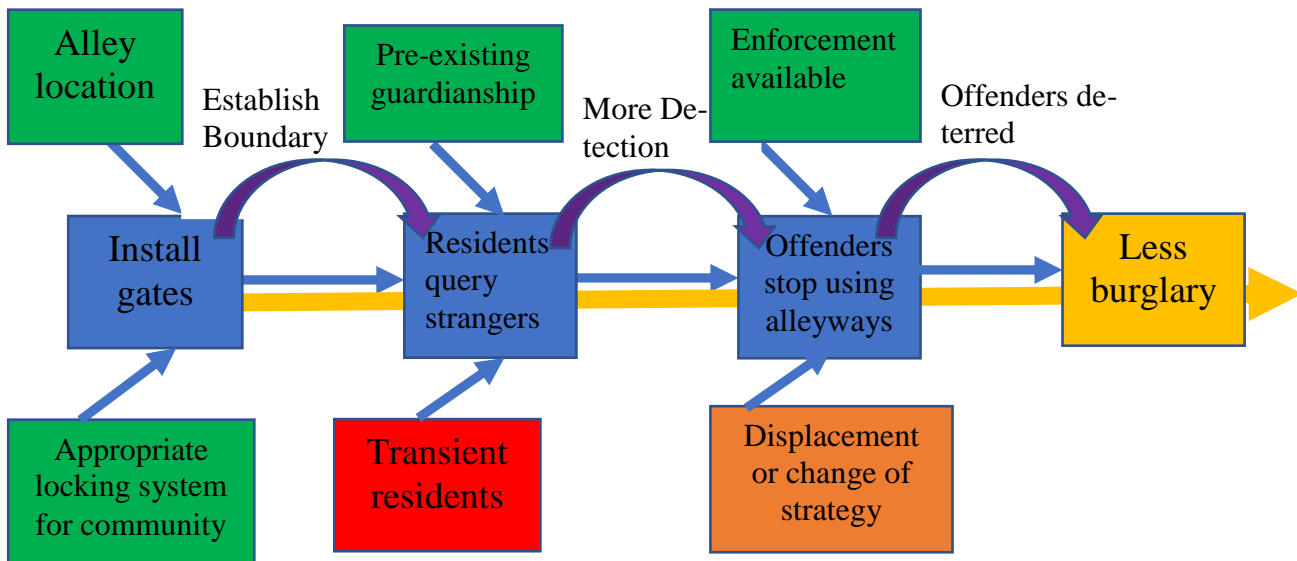


Figure 3: Example of a Street-level Theory of Change (STOC) for alley gating interventions

With this limited but suggestive evidence base, the authors then move on to what we see as more important for understanding what works for a police force that is considering introducing alley gates as an intervention. They discuss the mechanisms that are meant to help explain why the intervention works, the activities involved in implementation, moderating factors (what we call support factors), as well as potential barriers to effectiveness (derailers and off-setters in our language). They note that an important piece in the causal chain is some pre-existing notion of guardianship within the local area. People must be willing to challenge strangers in the alley or the gates will not have a significant effect. Their symbolic power to mark territory will be vitiated if the residents themselves do not treat that territory as restricted. A scheme for keeping track of who has keys and codes to the gates is important. The study suggests that student neighbourhoods, with a shifting transient population with little investment in the locality, may not be as promising a community for this sort of intervention as would be the case for a family-oriented neighbourhood.

It seems that the realist synthesis part of the study is carrying most of the weight of the analysis. The meta-analysis part is almost vestigial. The experimentally derived evidence tells you that installing gates worked to reduce burglary but that is only the beginning and end of a complex sequence of steps and interacting factors that we are interested in.

Indeed, there is a way in which a rigid insistence on prioritizing experimental evidence for *a particular intervention* may miss out on the strength of evidence that can be derived even from the experimental literature. Installing alley gates is an example of the more general intervention of

target hardening, in turn a subset of interventions aiming at 'situational crime prevention'. The activities of establishing or clarifying boundaries, using guardianship to detect more potential offenders and deterring offenders in such a way that they desist from committing crimes are broadly applicable. It is quite likely that contextual and moderating factors are more important for the success of each target hardening intervention than the precise form the target hardening takes (the alley gates in this case). For example, interventions upgrading doors and windows in areas targeted by burglars are premised on a similar causal framework. The study design insists on only looking at the specific intervention, thus excluding other experimental studies that may, in practice, be as relevant to demonstrating that (and how) alley gates will work.

Conclusion

In this chapter, we have introduced the origins, strengths and weaknesses of the contemporary medical model of evidence-based practice. We have shown how conventional evidence-based policing, although expanding the range of experimental research, continues to adhere to the notion that experimental research offers something uniquely compelling for establishing the causal effectiveness of interventions in any policy domain. Our argument is that such uniqueness is overemphasized because experimental evidence never stands alone when it comes to developing a causal model that might allow practitioners to learn when and where something is going to work, not just that it works somewhere.

Our conclusion is not that experimental approaches are without value, nor the motivation to explore and weight all rigorous sources of evidence that underlies the attraction of systematic reviews. But when it comes to practically exploring better policing strategies, a strong willingness to deviate from the medical model, especially the medical model of evidence-based practice, is better than fidelity to it. As with EBM+, EBPP+ has more help to offer. Our proposal for using STOCs to cast appropriate lights both on middle-range theories and support factors can boost this endeavour.

References

- Abramovaite J, Bandyopadhyay S and Dixon L (2015) The Dynamics of Intergenerational Family Abuse: A Focus on Child Maltreatment and Violence and Abuse in Intimate Relationships. *Journal of Interdisciplinary Economics* 27(2): 160–174. DOI: 10.1177/0260107915582254.
- Abramovaite J, Bandyopadhyay S, Bhattacharya S, et al. (2018) Alternatives to custody: Evidence from police force areas in England and Wales. *The British Journal of Criminology*. DOI: 10.1093/bjc/azy056.
- Alvarez-Dardet C and Ruiz MT (1993) Thomas McKeown and Archibald Cochrane: a journey through the diffusion of their ideas. *BMJ (Clinical research ed.)* 306(6887): 1252–1255.
- Bandyopadhyay S (2013) Crime Policy in an Era of Austerity. *The Police Journal* 86(2): 102–115. DOI: 10.1350/pojo.2013.86.2.617.
- Boles SM and Miotto K (2003) Substance abuse and violence. *Aggression and Violent Behavior* 8(2): 155–174. DOI: 10.1016/S1359-1789(01)00057-X.
- Bothwell LE, Greene JA, Podolsky SH, et al. (2016) Assessing the Gold Standard — Lessons from the History of RCTs. Malina D (ed.) *New England Journal of Medicine* 374(22): 2175–2181. DOI: 10.1056/NEJMms1604593.
- Bowers KJ, Johnson SD and Pease K (2005) Victimization and Re-victimisation Risk, Housing Type and Area: A Study of Interactions. *Crime Prevention and Community Safety* 7(1): 7–17. DOI: 10.1057/palgrave.cpcs.8140205.
- Braga AA, Welsh BC and Schnell C (2015) Can Policing Disorder Reduce Crime? A Systematic Review and Meta-analysis. *Journal of Research in Crime and Delinquency* 52(4): 567–588. DOI: 10.1177/0022427815576576.
- Cartwright N (1983) *How the Laws of Physics Lie*. Oxford University Press. DOI: 10.1093/0198247044.001.0001.
- Cartwright N (2007a) Are RCTs the Gold Standard? *BioSocieties* 2(1): 11–20. DOI: 10.1017/S1745855207005029.

- Cartwright N (2007b) *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge ; New York: Cambridge University Press.
- Cartwright N (2011) A philosopher's view of the long road from RCTs to effectiveness. *The Lancet* 377(9775): 1400–1401. DOI: 10.1016/S0140-6736(11)60563-1.
- Cartwright N (2013) Knowing what we are talking about: why evidence doesn't always travel. *Evidence & Policy: A Journal of Research, Debate and Practice* 9(1): 97–112. DOI: 10.1332/174426413X662581.
- Cartwright N and Cowen N (2014) *Making the Most of the Evidence in Education: A Guide for Working Out What Works Here and Now*. CHES Working Paper 2014–03. Durham University.
- Cartwright N and Hardie J (2012) *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford ; New York: Oxford University Press.
- Chalmers I (2008) Archie Cochrane (1909-1988). *Journal of the Royal Society of Medicine* 101(1): 41–44. DOI: 10.1258/jrsm.2007.071004.
- Cochrane AL (1972) *Effectiveness and Efficiency: Random Reflections on Health Services*. Nuffield Trust.
- Cook TD and Campbell DT (1979) *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin.
- Cowen N, Virk B, Mascarenhas-Keyes S, et al. (2017) Randomized Controlled Trials: How Can We Know “What Works”? *Critical Review* 29(3): 265–292. DOI: 10.1080/08913811.2017.1395223.
- Cullen FT and Gendreau P (2001) From Nothing Works to What Works: Changing Professional Ideology in the 21st Century. *The Prison Journal* 81(3): 313–338. DOI: 10.1177/0032885501081003002.
- Cutler D, Deaton A and Lleras-Muney A (2006) The Determinants of Mortality. *Journal of Economic Perspectives* 20(3): 97–120.

- Deaton A (2003) Health, Inequality, and Economic Development. *Journal of Economic Literature*: 46.
- Deaton A (2015) *The Great Escape: Health, Wealth, and the Origins of Inequality*. 6., printing, and 1. paperback printing. Princeton, NJ: Princeton Univ. Press.
- Deaton A and Cartwright N (2018) Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine* 210: 2–21. DOI: 10.1016/j.socscimed.2017.12.005.
- DeMets DL and Cook T (2018) Challenges of Non-Intention-to-Treat Analyses. *JAMA*. DOI: 10.1001/jama.2018.19192.
- Eisner M (2003) Long-Term Historical Trends in Violent Crime. *Crime and Justice* 30: 83–142. DOI: 10.1086/652229.
- Elster J (1998) A plea for mechanisms. In: Hedström P and Swedberg R (eds) *Social Mechanisms: An Analytical Approach to Social Theory*. Studies in Rationality and Social Change. Cambridge: Cambridge University Press, pp. 45–73. DOI: 10.1017/CBO9780511663901.003.
- Engel GL (1977) The need for a New Medical Model: A Challenge for Biomedicine. *Science* 196(4286): 129–136.
- Farrington DP and Petrosino A (2001) The Campbell Collaboration Crime and Justice Group. *Annals of the American Academy of Political and Social Science* 578: 15.
- Farrington DP, Gottfredson DC, Welsh BC, et al. (2002) Maryland Scientific Methods Scale. In: Sherman LW (ed.) *Evidence-Based Crime Prevention*. London ; New York: Routledge, pp. 13–21.
- Goldacre B (2015) Commentary: Randomized trials of controversial social interventions: slow progress in 50 years. *International Journal of Epidemiology* 44(1): 19–22. DOI: 10.1093/ije/dyv005.
- Guyatt G (1992) Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *JAMA* 268(17): 2420. DOI: 10.1001/jama.1992.03490170092032.

- Han L, Bandyopadhyay S and Bhattacharya S (2013) Determinants of violent and property crimes in England and Wales: a panel data analysis. *Applied Economics* 45(34): 4820–4830. DOI: 10.1080/00036846.2013.806782.
- Haynes L, Service O, Goldacre B, et al. (2012) *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*. London: Cabinet Office. Available at: <http://www.ssrn.com/abstract=2131581> (accessed 24 January 2019).
- Illich I (1982) *Medical Nemesis: The Expropriation of Health*. New York: Pantheon Books.
- Johnson SD, Summers L and Pease K (2009) Offender as Forager? A Direct Test of the Boost Account of Victimization. *Journal of Quantitative Criminology* 25(2): 181–200. DOI: 10.1007/s10940-008-9060-8.
- Johnson SD, Tilley N and Bowers KJ (2015) Introducing EMMIE: an evidence rating scale to encourage mixed-method crime prevention synthesis reviews. *Journal of Experimental Criminology* 11(3): 459–473. DOI: 10.1007/s11292-015-9238-7.
- Jones DS and Podolsky SH (2015) The history and fate of the gold standard. *The Lancet* 385(9977): 1502–1503. DOI: 10.1016/S0140-6736(15)60742-5.
- Joyce KE and Cartwright N (2018) Meeting our standards for educational justice: Doing our best with the evidence. *Theory and Research in Education* 16(1): 3–22. DOI: 10.1177/1477878518756565.
- Laing RD (1971) *The Politics of the Family and Other Essays*. London: Tavistock Publications.
- Lumsden K and Goode J (2018) Policing Research and the Rise of the ‘Evidence-Base’: Police Officer and Staff Understandings of Research, its Implementation and ‘What Works’. *Sociology* 52(4): 813–829. DOI: 10.1177/0038038516664684.
- Mackie JL (1965) Causes and Conditions. *American Philosophical Quarterly* 2(4): 245–264.
- Mackie JL (1980) *The Cement of the Universe*. Oxford University Press. DOI: 10.1093/0198246420.001.0001.

- Marchionni C and Reijula S (2018) What is mechanistic evidence, and why do we need it for evidence-based policy? *Studies in History and Philosophy of Science Part A*. DOI: 10.1016/j.shpsa.2018.08.003.
- Martinson R (1974) What Works? Questions and answers about prison reform. *The Public Interest* 35: 22–54.
- McKeown T (1976) *The Modern Rise of Population*. New York: Academic Press.
- Parkkinen V-P, Wallmann C, Wilde M, et al. (2018) *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*.
- Pawson R (2000) Middle-range realism. *European Journal of Sociology* 41(02): 283. DOI: 10.1017/S0003975600007050.
- Pawson R (2006) *Evidence-Based Policy: A Realist Perspective*. London ; Thousand Oaks, Calif: SAGE.
- Pawson R and Tilley N (1997) *Realistic Evaluation*. London ; Thousand Oaks, Calif: Sage.
- Petrosino A, Turpin-Petrosino C and Buehler J (2003) Scared Straight and Other Juvenile Awareness Programs for Preventing Juvenile Delinquency: A Systematic Review of the Randomized Experimental Evidence. *The ANNALS of the American Academy of Political and Social Science* 589(1): 41–62. DOI: 10.1177/0002716203254693.
- Rothman KJ (1976) CAUSES. *American Journal of Epidemiology* 104(6): 587–592. DOI: 10.1093/oxfordjournals.aje.a112335.
- Sackett DL, Rosenberg WMC, Gray JAM, et al. (1996) Evidence based medicine: what it is and what it isn't. *BMJ* 312(7023): 71–72. DOI: 10.1136/bmj.312.7023.71.
- Sherman L, Neyroud PW and Neyroud E (2016) The Cambridge Crime Harm Index: Measuring Total Harm from Crime Based on Sentencing Guidelines. *Policing* 10(3): 171–183. DOI: 10.1093/police/paw003.
- Sherman LW (ed.) (1997) *Preventing Crime: What Works, What Doesn't, What's Promising: A Report to the United States Congress*. Office of Justice Programs research report. College Park, MD: Univ. of Maryland.

- Sidebottom A, Tompson L, Thornton A, et al. (2017) *GATING ALLEYS TO REDUCE CRIME: A METAANALYSIS AND REALIST SYNTHESIS*. WHAT WORKS: CRIME REDUCTION SYSTEMATIC REVIEW SERIES 1, August. Coventry: College of Policing.
- Smith GCS (2003) Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ* 327(7429): 1459–1461. DOI: 10.1136/bmj.327.7429.1459.
- Starr M, Chalmers I, Clarke M, et al. (2009) The origins, evolution, and future of The Cochrane Database of Systematic Reviews. *International Journal of Technology Assessment in Health Care* 25(S1): 182–195. DOI: 10.1017/S026646230909062X.
- van Mastrigt SB and Farrington DP (2009) Co-Offending, Age, Gender and Crime Type: Implications for Criminal Justice Policy. *British Journal of Criminology* 49(4): 552–573. DOI: 10.1093/bjc/azp021.
- Weisburd D, Wyckoff LA, Ready J, et al. (2006) Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. *Criminology* 44(3): 549–592.
- Weisburd D, Hinkle JC, Braga AA, et al. (2015) Understanding the Mechanisms Underlying Broken Windows Policing: The Need for Evaluation Evidence. *Journal of Research in Crime and Delinquency* 52(4): 589–608. DOI: 10.1177/0022427815577837.
- Weisburd D, Farrington DP, Gill C, et al. (2017) What Works in Crime Prevention and Rehabilitation: An Assessment of Systematic Reviews. *Criminology & Public Policy* 16(2): 415–449. DOI: 10.1111/1745-9133.12298.
- Welsh BC and Farrington DP (2001) Toward an Evidence-Based Approach to Preventing Crime. *The ANNALS of the American Academy of Political and Social Science* 578(1): 158–173. DOI: 10.1177/000271620157800110.
- Wilson DB (2001) Meta-Analytic Methods for Criminology. *The ANNALS of the American Academy of Political and Social Science* 578(1): 71–89. DOI: 10.1177/000271620157800105.
- Witt R, Clarke A and Fielding N (1999) Crime and Economic Activity: a panel data approach. *British Journal of Criminology* 39(3): 391–400.