The Future of Evidence-Based Policing

Cambridge University Press

David Weisburd

Tal Jonathan-Zamir

Badi Hasisi

Gali Perry

CHAPTER 15:

The Role of Randomized Experiments in Developing the Evidence for Evidence-Based Policing

By

Lorraine Mazerolle[*]

Elizabeth Eggins[*]

Lorelei Hine[*]

Angela Higginson[#]

* The University of Queensland

Queensland University of Technology

Acknowledgements*: Development of the GPD was funded by the Australian Research Council and the College of Policing (United Kingdom). Funding to update the GPD has been provided by grants for systematic reviews using the GPD, from sources including Arnold Ventures, the Australian Institute of Criminology and the Department of Homeland Security (United States). We also acknowledge the 156 volunteer student interns who have contributed over 7,000 hours of work on the GPD since 2015.*

Abstract

Randomized experiments, quasi-experiments and systematic reviews are critical for developing the evidence for evidence-based policing. In this paper, we explore the role of experiments in contributing to the evidence base in policing. Drawing on the Global Policing Database (GPD) corpus of 3,487 high quality evaluation studies in policing, including 431 randomized controlled trials, we describe the breadth and depth of the evidence base in policing. We find that randomized controlled trials form only about 12 percent of the total evidence based in policing yet this small number of trials has had enormous policy influence over the last fifty years. We also find that most of the evidence in policing is around frontline policing practices and about half of the RCTs in the world come from the US: a greater proportion of RCTs coming from outside of the US than previously reported. We conclude that the breadth of people and places generating high quality evidence will help generalize policing policies and practices beyond the US and is likely to have a snowball effect in fostering the next generation of experimentalists in policing.

Introduction

An important debate around the Evidence-Based Policing movement is determining what constitutes "evidence" and what does not. Sherman's original rendition of evidence-based policing (see Sherman, 1998) argued for a strong foundation of scientific evidence to guide proactive efforts to underpin policing practices. Sherman (1998) based these arguments on the medical model which prioritized randomized control trials (RCT) as the primary source of research "evidence." Over the years this definition of what constitutes scientific evidence has led to considerable debate. Practitioners often value their experiences as evidence, rewarding knowledge of the standard operating procedure and "hunches" in investigations over the input from academics (Lum & Koper, 2017). Policy-makers draw on political interpretation and social norms to influence their interpretation and acceptance of research (Head, 2008). Academics are similarly divided regarding what evidence is and how it influences policy and practice. Lum and colleagues (2006), for example, prioritize systematic reviews and experimental studies in shaping policy and practice responses to crime problems (Lum & Kennedy, 2012). Laycock (2012), on the other hand, argues that there is great value in adopting a broad concept of what constitutes evidence, drawing from a wide range of methodological approaches to understand what works in policing.

Numerous initiatives over the last decade have sought to gather together the scientific evidence on policing – with "evidence" defined in different ways – to make information about policing practices readily and openly available to researchers, policy makers and practitioners. At least six open access websites offer evidence on policing practices, each compiling the research evidence with different search methods, different scoping of outcome measures and different evidence inclusion criteria. For example, the Campbell Collaboration (www.campbellcollaboration.org), the Evidence Based Policing Matrix

(http://cebcp.org/evidence-based-policing/the-matrix/), Crime Solutions (http://www.crimesolutions.gov), the Crime Reduction Toolkit (http://whatworks.college.police.uk/toolkit/Pages/Toolkit.aspx), the Center for Problem-Oriented Policing (www.popcenter.org) and the Global Policing Database (GPD, www.gpd.uq.edu.au) all provide policy makers and practitioners with easy access to different types of research evidence about policing practices. Collectively these websites offer police comprehensive information about what is known about the effectiveness of police practices and tread a fine line between satisfying the desire for the highest quality scientific "evidence," yet at the same time providing systematic access to a substantial breadth of policing research generated across the world.

We begin this chapter with an historical examination of the role of randomized experiments in developing the scientific evidence for evidence-based policing. We then use the GPD to show the trends in the use of randomized experimental studies in policing. We identify the types of programs more likely to be evaluated using randomized experiments and conclude with a discussion of the significance of experiments in shaping the global evidence base in policing.

Experimentation in Policing

Anthony Braga and colleagues (2014), in their study of randomized experiments in policing, track the maturing of the experimental policing movement since the 1970s. They argue that, by the turn of the century, there was sufficient critical mass in human and social capital to "…shift scientific paradigms in mainstream criminology towards increased experimentation" (Braga et al., 2014: 22). In this section, we explore the key milestones and turning points that

have characterized this process of maturing in the use of experiments in policing, providing the background needed to understand the current status of experimental evidence in policing.

The story of police experimentation really begins with the Police Foundation in the early 1970s. The Foundation was established by a grant from the Ford Foundation as a non-profit, non-partisan and independent research hub that sought to advance policing through innovation and science. Over the last fifty years a parade of highly respected police practitioner leaders (including James Q Wilson, Patrick V Murphy, Hubert Williams and Jim Bueermann) and police scientists (including Lawrence Sherman, George Kelling and David Weisburd) have served in a variety of different capacities for the Police Foundation, conducting some of the most influential experiments in the field of policing.

One of the most widely cited experiments undertaken by the Police Foundation is the 1970 Kansas City Preventive Patrol Experiment. Led by George Kelling and his colleagues Pate, Dieckman and Brown (see Kelling et al., 1974), the experiment found that random patrol had very little impact on crime and community perceptions of safety, challenging the prevailing wisdom at the time that assumed preventive patrols were effective at deterring crime. The Kansas City experiment set the scene for a number of subsequent Police Foundation experiments that were fundamental for shaping new thinking in police policy and practice across the world. For example, the Newark Foot Patrol Experiment found that fair engagement between police and community residents reduced fear of crime (Kelling et al., 1981). The experiences and results from the Newark Foot Patrol Experiment were significant in shaping the highly influential Broken Windows thesis, proposed by Police Foundation Chair James Q Wilson and foundation-based scientist George Kelling. The original Broken Windows paper published in the Atlantic Monthly (Kelling & Wilson, 1982) is identified as

one of the most cited and popular articles ever published in The Atlantic Monthly in its 163-year history.

By the mid-1980s, the preventive value of police started to receive renewed attention leading to a second wave of policing experiments that have again shaped and defined policy and practice across the globe. In 1988, Lawrence Sherman and David Weisburd received funding from the National Institute of Justice to conduct the Minneapolis Hot Spots Experiment to test whether or not directed patrols in high crime places could reduce crime (see Sherman and Weisburd, 1995). Sherman and Weisburd (1995) used what had been recently discovered about the consistent patterns of crime clustering in very small micro places (see Sherman, Gartin, & Buerger, 1989; see also Pierce, Spaar, & Briggs, 1988) to conceptualize crime hotspots as the experimental units of analysis. With random allocation of 110 crime hot spots (55 receiving, on average, twice as much observed patrol presence as the 55 control hot spots), they discovered that substantial increases in police patrol presence could lead to reductions in crime and disorder in crime hotspots (Sherman & Weisburd, 1995). This experiment was the starting point for the widespread adoption of directed patrol policies targeted at crime hotspots across the world (see National Academies of Sciences report, 2018).

Braga and colleagues' (2014) study of networks of police scholars examines the people who have shaped the use of experimentation in policing. Their social network analysis shows that Weisburd and Sherman, both mentored by Albert J. Reiss, Jr. when they were doctoral students in sociology at Yale University, "form two highly central nodes that have proliferated nearly the entire social network of experiments in policing undertaken up until 2011" (see Braga et al., 2014: 17). The impact of the experimental work of Sherman and

Weisburd is significant and most recently acknowledged in them ranking numbers 1 and 2 as the world's most influential scholars in criminology (https://academicinfluence.com/articles/people/most-influential-criminologists-today). For two experimental criminologists and police scholars to occupy those positions of influence is indicative of the role of field experimentation in shaping public policy, scholarly thinking and policing practice.

Beyond the influence of individual policing experiments, systematic reviews of the research evidence are highly reliant on a healthy supply chain of experiments in policing. The establishment of the Campbell Collaboration in 2000 (see Farrington & Petrosino, 2001) with the Crime and Justice Group as a foundation coordinating group, subsequently set up a policy appetite and demand for high quality evaluations of policing practices. At the time of writing, 26 systematic reviews in policing have been undertaken under the auspices of the Campbell Collaboration, collectively generating more than 3,635 citations. The policy impacts of these Campbell policing reviews are extensive. For example, the Braga and colleagues (2007; updated 2019) review of hotspots policing has been a key driver behind reforms in policing strategies across the world. In Trinidad and Tobago the Braga et al. (2007) review was the foundation for a decision to implement hot spots patrols to tackle a serious escalation in gun-related homicide. Commissioner Williams of Trinidad and Tobago started the national roll out of hot spots patrols in 2014 following promising results from a randomized controlled trial undertaken locally. The new approach to policing patrols has shown a significant impact on the level of gun-related homicide (Sherman et al., 2014). In Colombia, the Braga et al. review (2007) led to discussions with police in Medellin and Bogota, resulting in RCTs on hot spots policing in both cities (Collazos, Garcia, Mejia, Ortega, & Tobón, 2020; Blattman, Green, Ortega, & Tobón, 2018).

Despite the clear influence of experimentation in policing over the last fifty years, randomized field trials that test policing practices comprise just a small fraction of the research evidence in policing (see also Garner & Visher, 2003). Braga and colleagues' (2014) systematic search of policing experiments find a total of just 63 policing experiments ever undertaken from 1970 to 2011, with the pace of experimentation starting to increase in the mid to late 1980s and then proliferating during the 1990s. Interestingly, 75 percent of all of the policing experiments identified in the 2011 search were conducted in the United States. Neyroud (2017) in his search of policing experiments conducted up to and including 2017 finds a total of 122 randomized controlled trials in policing. What both the Braga et al (2014) and Neyroud (2017) studies suggest is that police experimentation is highly influential yet a relatively rare method for shaping the policing evidence base. In the sections that follow we take a systematic approach to updating these prior policing experiment searches and exploring the number, scope and influence of experimentation in policing in shaping the global evidence base.

The Global Policing Database

To better understand and quantify the role of experimentation in policing, we draw from the Global Policing Database (GPD). The GPD is a web-based exhaustive repository of intervention research relating to police and policing practices. The GPD is designed to capture all high quality published and unpublished impact evaluations of interventions relating to police and policing conducted since 1950. Using systematic search and review techniques, the GPD compiles all experimental and quasi-experimental evaluations into one searchable location. It places no limits on outcome measure(s) and includes any policing intervention approach with no restrictions on the type of policing techniques or language. The

GPD aims to be updated at least biennially, facilitating an up-to-date body of work from which practitioners, policy makers, academics and those interested in policing can derive knowledge and evidence. The goal of the GPD repository is to encourage the translation of research into evidence-based practices, thereby cultivating a global approach to policing practice.

In the GPD, evidence is captured through a series of comprehensive systematic searches for documents that are about police and policing. Indeed, 293,616 unique records from 1950 to 2018 constitute the corpus of studies that are captured in the GPD search prior to systematic screening for evaluation studies. Of the documents screened to date, a total of 3,586 documents are identified as impact evaluation studies using evaluation methods that meet our criteria[1], representing 2.2% of the N=165,525 documents that were screened at title and abstract screening stage, 6% of the N=59,927 documents that were screened at Stage 1 full-text screening stage and 23.4% of the N=15,339 documents that were screened at Stage 2 full -text screening stage. A summary of the systematic screening stages of the GPD process are summarised in Figure 1.

< Insert Figure 1 here >

Types of Evaluation Designs

The protocol for the GPD describes the types of evaluation designs that are eligible for inclusion in the GPD corpus. These are designs that are contain a quantitative impact evaluation of policing interventions that, at a minimum, include a valid comparison group.

[1] Please refer to the Global Policing Database Protocol that describes the search and multiple stages of screening. (https://gpd.uq.edu.au/files/original/86fc6980ee633bce119287327d4a432b646dd515.pdf).

The GPD includes studies where the comparison group receives no intervention, 'business-as-usual' policing or an alternative type of intervention[2]. At the time of writing there is a caveat with reporting on the status of the contents of the GPD because our first and second stage screening of documents identified in the systematic search are prioritized in two specific ways: first, by date, with the most recent records being screened first before moving backwards in time and second, by the specific search terms that have been used by researchers utilizing the GPD for specific systematic reviews. Final eligibility screening – the most time and training intensive component of the GPD compilation – has been targeted to documents that expedite specific systematic reviews. As such, the current corpus of eligible evaluations in the GPD are slightly skewed toward specific topics that have been the subject of GPD-related research grants and contracts.

At the time of writing, the GPD has been used to produce eight Campbell Collaboration reviews including Body Worn Cameras (see Lum et al., 2020) and police programs to increase community connectedness (see Mazerolle et al., 2020); updates of three Campbell reviews including problem-oriented policing (POP) (see Hinkle et al., 2020), legitimacy policing (Mazerolle, Bennett, et al., 2018) and street level drug law enforcement (see Mazerolle, Eggins, et al., 2020); three in-progress Campbell Collaboration reviews including Third Party Policing (Mazerolle et al., 2016), multiagency interventions with police as a partner for preventing radicalisation to violence review (Mazerolle, Cherney et al., 2020) and

[2] The list of research designs included in the GPD are as follows: systematic reviews with or without meta-analyses; cross-over designs; cost-benefit analyses; regression discontinuity designs; cost-benefit analyses; regression discontinuity designs; designs using multivariate controls (e.g., multiple regression); matched control group designs with or without pre-intervention baseline measures (propensity or statistically matched); unmatched control group designs with pre-post intervention measures which allow for difference-in-difference analysis; unmatched control group designs without pre-intervention measures where the control group has face validity; short interrupted time-series designs with control group (less than 25 pre- and 25 post-intervention observations); long interrupted time-series designs with or without a control group (≥25 pre- and post-intervention observations; raw unadjusted correlational designs where the variation in the level of the intervention is compared to the variation in the level of the outcome.

community-oriented policing to reduce crime, disorder, and fear and improve legitimacy and satisfaction with police (Gill et al., 2017); thee non-Campbell Collaboration reviews including police techniques for investigating serious violent crime (Higginson et al., 2017), criminal justice responses to Child Exploitation Material (CEM) offending (Eggins, Mazerolle, et al., 2020), and the impact of supplier arrests and seizures on drug crime, drug use, drug price, drug purity, and drug harm outcomes (Eggins, Higginson et al., 2020). Research teams have also used the GPD to conduct six rapid reviews on a variety of policing topics including police partnerships with agencies to tackle mental health problems (Eggins, Mazerolle et al., 2020), child sex offender risk management (Eggins, Mazerolle, Hine et al., 2020), organizational approaches to road policing (Mazerolle, Eggins, et al., 2019b), best practices in policing alcohol impaired driving (Mazerolle, Eggins, et al., 2019a), police training (Bennett & Newman, 2018) and criminal justice responses to domestic violence, including a review of policing domestic violence (Mazerolle, Eggins, Sydes, et al., 2019; Mazerolle, Eggins, et al., 2018). All of these review topics have skewed the distribution of screened records in the GPD which influences, to some degree, the data presented in this paper in terms of the trends by year and by policing practice. Figure 2 provides a pie chart illustrating the breakdown of screened and eligible studies in the GPD by RCT, systematic review and quasi-experimental design.

< Insert Figure 2 here >

As this chart shows, the majority of the studies in the GPD (of the 3,586) are quasi-experimental designs (82%), with RCTs comprising only 12 percent (N = 431) of all documents systematically processed to date according to the GPD protocol. Six percent of the studies in the GPD are systematic reviews with or without meta-analysis, presenting a high

quality of synthesised material. The GPD excludes single group designs with pre- and post-intervention measures as these designs are highly subject to bias and threats to internal validity.

Compared to the earlier searches by Braga and colleagues (2014) and Neyroud (2017), the number of experiments in policing identified in the GPD is far greater than what had previously been uncovered. In our GPD systematic search for policing evidence, we identify 431 policing experiments across the world (at the time of writing). This is a 584% increase on the number of policing experiments located by Braga et al. (2014) and 253% increase over what was found by Neyroud (2017). Importantly, this increase in located policing experiments are not just from the years post the searches by Braga and colleagues (2014) and Neyroud (2017). Rather, our revisiting of the years 2000 to the present time has located previously undiscovered policing experiments using the GPD search methods. We explore the nature of these newly located experiments further in the sections below.

When we examine the trends in police experimentation over time, our analysis of the GPD data confirms Braga and colleagues (2014) earlier finding that experimentation had waves of increasing popularity in the 1980s and then again in the 1990s. Figure 3 illustrates the type of designs the GPD systematic search has identified over time. Over the 15 year period from 2004 to 2018 there is a slight upward trend in the quantity of studies identified by the GPD systematic search. The only exception is a down-turn in 2009, which is presumably the effects of the global financial crisis on funding for research. The quantities of studies which use an RCT or are systematic reviews, with or without meta-anlysis, show a steady upward trajectory. This indicates that these designs, specifically RCTs, are gaining traction amongst

researchers and becoming more common in shaping the body of knowledge relating to police and policing practitices.

< Insert Figure 3 here >

The increase in systematic reviews, albeit small, is perhaps in response to the creation of the Campbell Collaboration in 2000 and the funding to support a cluster of reviews by the UK National Policing Improvement Agency (now part of the UK College of Policing) in 2008 (see Wilson et al., 2020). Figure 3 also shows a general upward trend in studies using a quasi-experimental design, with some year-to-year fluctuations. But what is of primary interest to this paper, Figure 3 shows the number of RCTs per year have increased each and every year since the turn of the century: the calendar year 2005 shows just 6 new RCTs in policing published, in 2010, there were 28 studies published and then in 2013 there were 44 policing RCTs just in that calendar year and then in 2017 there were 55 which is a 817% increase in the number of policing RCTs produced from 2005 to 2017.

There are a range of reasons why we see this extraordinary increase in the number of RCTs in policing. First, the efforts of the field RCT pioneers in policing, led by Sherman and Weisburd, demonstrated that field experimentation in policing could be done to a high level of scientific rigor. This laid the foundations for an increase in the confidence of RCTs to be fair tests of policing practices. Second, the highly influential Maryland Report (see Sherman et al., 1997) captured the imagination of policy advocates, most prominently Jon Baron and Jerry Lee, who invested heavily in the establishment of centres of experimental excellence both at the University of Pennsylvania and then later at Cambridge University under the leadership of Lawrence Sherman. Jerry Lee in particular was instrumental in pushing the

global need for RCTs to inform criminal justice policy and practice. He went on to become a major sponsor of the Stockholm Prize in Criminology that awarded the prize based on strong scientific methods in the discovery and use of evidence to shape global crime and justice policy and practice. Third, the establishment of the Campbell Collaboration in 2000 (see Farrington & Petrosino, 2001) created an unprecedented demand for RCTs in particular and high quality evaluations of crime and justice practices, including policing practices. Fourth, funding agencies such as the National Institute of Justice in the US, the Australian Research Council and the UK Economic and Social Research Council began funding RCTs at a rate not previously seen at the end of the 20th century. Telep and his colleagues (2015) offer some insights around the increase in funding experimentation. Using data from official records of grant awards made by the US National Institute of Justice between fiscal years 2001 and 2013, they categorized awards based on study design (including RCT design). They identified 99 grant awards for experiments where support for the use of experimental designs increased during their 13-year study period, dwarfing the funding support for the use of experimental designs providing during the 1990s (Telep et al., 2015).

The GPD repository of policing evidence also challenges the earlier thinking that the vast majority of experiments had been proliferated by "the salient few" (see Braga et al., 2014). Of the 431 experiments in the GPD, just 37 of these experiments (8.6% percent) have been undertaken by scholars listed as lead experimental policing scholars in the Braga et al. (2014) study (see Table 2, page 13[3]). This suggests that there are likely a much larger number of scholars throughout the world who are producing police experiments that have previously gone unnoticed in prior audits of policing experiments. The emergence and escalating pace

[3] These scholars include Sherman, Strang, Davis, Weisburd, Pate, Mazerolle, Angel, Taylor, Braga, Dunford and McCold (see Table 2, page 13).

over the last twenty years of systematic reviews in policing will bring to light many of these previously hidden experiments and high quality quasi experiments in policing. Our discovery of a treasure trove of policing experimentation is a boost to capacity of researchers to conduct many more systematic reviews of policing practice and thereby increase the role of scientific evidence in shaping policy and practice.

Categories of Police and Policing Studies

The GPD can be used to examine the range of different topics of police experimentation. To be included in the GPD, each document must contain an impact evaluation of a policing intervention. Policing interventions are defined in the GPD as some kind of a strategy, program, technique, approach, activity, campaign, training, directive, or funding/organisational change that involves police in some way (other agencies or organisations can be involved). Police involvement is broadly defined as police initiation, development or leadership; or police as recipients of the intervention; or the intervention is related, focused or targeted to police practices; or that there is delivery or implementation of the intervention by police. To gain a better understanding of the type of interventions being implemented and tested, the GPD developed categories for the policing interventions. These cover **five** broad areas which distinguishes the intervention based on the role of police in the intervention.

The first category examines **policing practices with police as implementers**. These interventions examine policing practices such as hot-spot policing (the targeting of activities and resources to the areas where crime is most concentrated), problem oriented policing (POP; the identification and analysis of a specific crime problem to develop a targeted response) and order-maintenance policing (the management of minor offences and

neighbourhood disorder, in line with broken windows theory, to reduce community level crime). An example that falls into this category is Braga and Bond's (2008) experiment that evaluated the effects of policing disorder in Lowell, Massachusetts is categorized as a study where the police are the implementers of the intervention. In this study, the officers implemented a general policing disorder strategy finding that the strongest crime-prevention gains were generated by situational prevention strategies.

The policing practices category also includes **police as intervention partners**, where the police seek to co-produce crime control by working with citizens, other public sectors or private industry. Some of the interventions that are in this category include third-party policing (where police persuade or coerce third parties, such as businesses or other government agencies, to partner with police and activate their legal levers to assist in crime control) and community policing (where police seek to partner with citizens to promote crime control within communities, such as neighbourhood watch). For example, Mazerolle and her colleagues (2017) evaluated, under randomized field trial conditions, the deterrent effects of a police–school partnership, called the Ability School Engagement Program (ASEP). The partnership with schools sought to co-produce truancy reduction by actively engaging parents and their truanting children in a group conference dialogue that was designed to increase parental and child awareness of the truancy laws (and the consequences of noncompliance), and thereby foster students' willingness to attend school. The authors found that the police–school partnership intervention increased parental awareness of prosecution likelihood, which moderated students' self-reported willingness to attend school (Mazerolle et al., 2017).

The second category of studies in the GPD are **investigative techniques** which includes interrogation and forensic evidence gathering. Woody, Forrest and Yendra (2014), for

example, compared the effects of explicit and implicit false-evidence ploys on mock jurors' verdicts, sentencing recommendations, and perceptions of police interrogation. They explored how false-evidence ploys (FEP) in police interrogations were viewed by jurors, comparing implicit FEP, where police interrogators ask the suspects about potential evidence without making a claim of possessing the evidence, and explicit FEP, which involve direct claims from by interrogators about false evidence. Their experiment involved 255 college students randomly assigned to one of three conditions; implicit FEP, explicit FEP or control with results indicating that jurors do not discriminate between the FEP methods, therefore causing implications in how police use such methods in suspect interrogations (Woody et al., 2014).

The third category contains studies examining **police organisational structure, department sizes, and human resource policies**. An example in the GPD of a study in this category is by Ater, Givati and Rigbi (2014) who examined how the organizational structure of law enforcement agencies affected police activity and crime. They published their results in the *Journal of Public Economics*, examining the consequences of an organizational reform in Israel that transferred the responsibility for housing arrestees from the police to the prison authority. Their evaluation framework capitalized on a staggered rollout of the reform in different regions of the country, showing how organizational change led to an increase of 11% in the number of arrests and to a decrease of 4% in the number of reported crimes.

The fourth category covers interventions which **test police tools, techniques and technologies**. This category of studies includes Body Worn Cameras, Tasers, forensic techniques (such as collection and processing), line-up practices, and investigative interview protocols. For example, in an RCT by Sousa, Ready and Ault (2010), they present findings from a randomized field-training experiment designed to study the impact TASERs on police

officers' use-of-force decisions. In this study, officers were randomly assigned to either a treatment group (with TASERs) or a control group (without TASERs) to assess the extent that officers armed with the TASER use it as an alternative to other types of less-lethal force, finding that officers with TASERs are less likely to deploy pepper spray and the baton in response to aggressive physical resistance.

The fifth category includes **legislative, regulatory or policy reforms or changes affecting police**. The GPD includes a range of studies from different countries around the world that fall into this category yet none, for obvious reasons, use RCTs to evaluate legislative reform. For example, in a study by Heide and colleagues (2012) on police custody regulations, they analysed 3,674 medical records on fitness for custody, taken from two large German towns (Halle and Bremen). They found that the introduction of new police custody regulations in Halle had a significant influence on the medical decision on fitness for custody.

Table 1 below presents the different intervention categories by study design coded only for a sub-set of N=1,149 documents that have been used in reviews we have conducted so far (see earlier for the list of different review topics that have driven the focus of screening within the GPD). As Table 1 shows, from this selective sample of Stage 2 screened and eligible documents in the GPD, we find 181 RCTs (16 percent, which is a slightly higher percent than for the whole of the GPD), 63 systematic reviews (5.5 percent) and 905 quasi-experiments in policing (78.8 percent, which is slightly less than for the whole GPD), for a total of 1,149 studies that have been Stage 2 screened and uploaded to the GPD website[4].

[4] We note that there are many more studies in the GPD at various stages of processing. As such, this represents a somewhat biased sample of GPD cases that are weighted to the systematic searches we have conducted so far using the GPD.

What we find from this analysis is that policing practices comprise the largest category of policing studies using RCTs (N = 67), with 60 RCTs in policing that relate to investigative techniques, including interrogation and different forms of evidence gathering and 39 RCTs in police organizational reform and training. Table 1 also shows a total of 63 systematic reviews in policing (at least in this Stage 2 of GPD screening), most of which (N = 44) being about policing practices such as hotspots policing, problem oriented policing, pulling levers and community policing. This sample of studies also contain 905 quasi-experimental studies, with N = 378 being about policing practices and 335 being about police organizational reform, staffing and training.

< Insert Table 1 here >

The fact that the GPD reveals that the majority of policing experiments have added to the evidence base in the category of policing practices -- as opposed to policing tools or technologies, or organizational reforms or HR practices, or legislative reforms, or investigative techniques – speaks volumes about the topic focus for policing researchers and funding agencies. It suggests that funding agencies are highly willing to fund researchers to better understand what policing practices (like POP, hotspots policing, co-responder models, community policing, pulling levers) actually work to reduce crime and disorder. Given that policing practices comprise a significant portion of the dollar value attached to the deployment of police at the frontline, a strong evidence base in policing practices is an important foundation for future rationalization and focus in the way scarce policing resources are deployed.

Global Contributions to the Evidence on Policing

One of the important questions about the evidence base in policing is whether or not the research is relevant to just a small number of countries or whether it has global relevance. Braga and his colleagues (2014) identified 75% of the world's contribution to policing RCTs to have been undertaken in the US suggesting significant limitations in the generalizability of the body of RCT evidence beyond the US context. Our GPD systematic search and review of the policing evidence based reveals a slightly different story where 50.1% (n = 216) of all the RCTs identified in the GPD have been conducted in the US. Table 2 presents the range of countries contributing quasi experimental, systematic review and RCT studies to the GPD between 2004 and 2018.

< Insert Table 2 here >

As Table 2 shows, the majority (n = 1822; 52.3%) of all GPD eligible studies (quasi, RCT and systematic reviews) originate from the United States. The top 20 countries see a high prevalence of European work, namely from the United Kingdom (n = 241; 6.9%). Australia also has a high contribution to GPD with 190 (5.4%) of the studies in the GPD. Of the top 10 countries, India (n = 42; 1.2%) is the only one of non-American, European or Oceanic status. Countries from Africa, Asia, South America and the Middle East appear in the top 20, although are under-represented with less than 1% of the publications per country. Comparative research on policing comprises 7.1 percent of the total research. In total, 87 different countries have contributed studies to the GPD.

When we examine the global contributions to the population of randomized controlled experiments in policing, Table 2 shows that 216 (50.1%) of all experiments were conducted

in the US. This tells a slightly different story to what was previously reported in Braga et al (2014) where 75% of all policing experiments were reported as stemming from the US. There are several reasons for why we find a greater global distribution of experimental contributions to the evidence base. First, our search is more recent than the Braga et al (2014) search. Second, our search is far more comprehensive than previous searches, with more than 89 grey literature search locations as well as 88 databases accessed to build the corpus of evidence in the GPD. Many of the grey literature sources and accessed databases are non-criminal justice databases and are not biased towards US sources, helping to cast a much wider net than was undertaken by Braga and his colleagues (2017) or by Neyroud (2017). Third, our outcome criteria for inclusion in the GPD is broader than previous searches where we have included studies examining a range of outcomes such as police physical or psychosocial wellbeing, community perceptions (including fear of crime) and efficiency outcomes of HR practices, such as shift rosters. Overall, whilst the US still contributes about half of the worlds' RCT evidence, the GPD reveals that there are at least 32 other countries that have contributed RCTs to the evidence base. This is also a likely underestimation of the global reach of policing experimentation given the biases in the GPD to English written documents.

Discussion and Conclusion

Randomized experiments, quasi-experiments and systematic reviews are critical for developing the evidence for evidence-based policing. In this paper, we explored the role of experiments in contributing to the evidence base in policing. Drawing on the Global Policing Database (GPD) with the full corpus of studies at the time of writing (June 2020), we describe the breadth and depth of high quality scientific evidence, particularly focusing on RCTs, that form the foundation of the evidence base in policing.

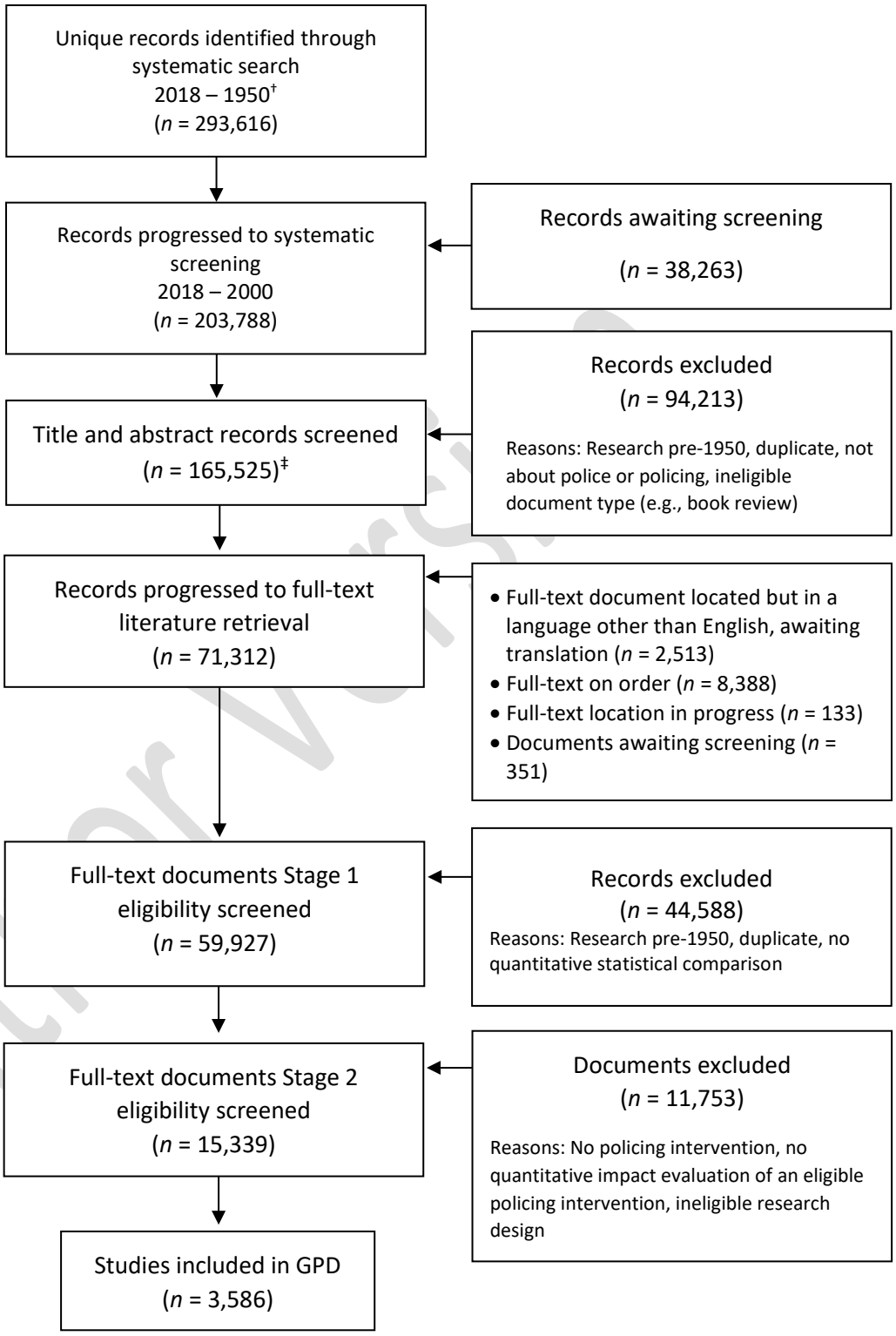
Our exploration of the population of RCTs in policing offers three primary insights: first, randomized controlled trials form only about 12 percent of the total evidence base in policing, reviews contribute about 6 percent and quasi-experimental designs about 82 percent, suggesting the need for decision makers to draw from a variety of evaluation designs – yet still strong designs - in formulating policy and practice decisions. We conclude, however, that the relatively small number of RCTs have punched above their weight in terms of policy influence.

Second, our review of the evidence from the GPD suggests a much broader group of scholars and countries contributing RCT evidence than was previously reported (see Braga et al, 2014). We find that about half of the RCTs identified by the GPD come from the US and fewer than 10% are generated by what Braga and colleagues (2014) referred to as the "salient few." This is a good news story for the RCT evidence base as it means that more countries and more scholars across the world are now contributing to the policing evidence base than previously thought. This breadth of people and places generating high quality evidence will help generalize policing policies and practices beyond the US and is likely to have a snowball effect in fostering the next generation of experimentalists in policing.

Finally, our study finds that most of the evidence in policing is around frontline policing practices (such as hotspots policing, problem oriented policing, third party policing, community policing). Other policing strategies, programs, techniques, approaches, activities, campaigns, training, directive, or organisational change involving police were found in the GPD, but mainly as quasi-experimental studies. These other types of policing interventions include investigative techniques; studies examining police organisational structure,

department sizes, and human resource policies; interventions which test police tools, techniques and technologies; and legislative, regulatory or policy reforms or changes affecting police. We argue, therefore, that the use of RCTs to inform policy and practice will need to be developed more broadly across the business of policing beyond frontline service delivery to facilitate policing to become more evidence-based into the future.

Figure 1:
Prisma

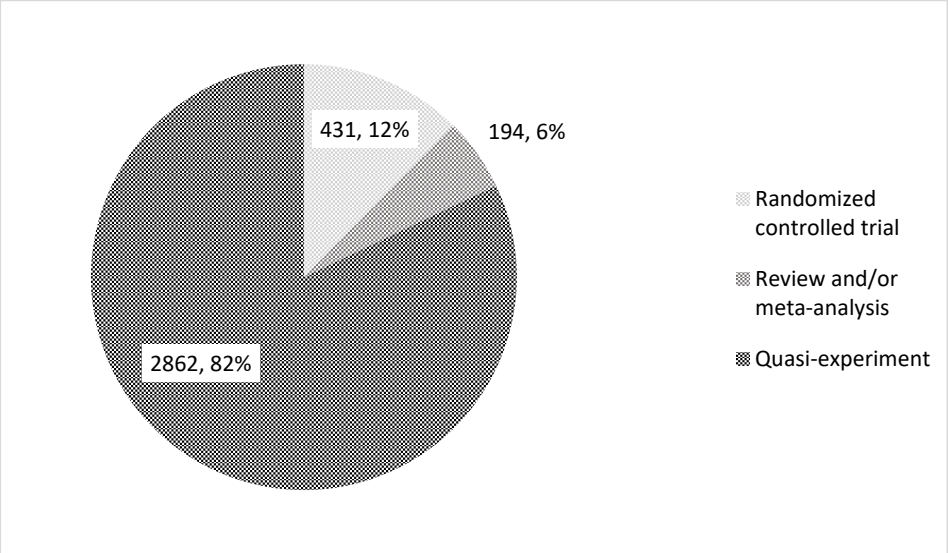


† Search period 1 January 1950 - 31 December 2018. Number represents N raw search results which have had duplicates and clearly ineligible document types removed (e.g., book reviews, tables of contents). Estimate only, as data cleaning for <2000 still in progress.

‡Represents 2018 – 2000 (≈81.23% of all records between 2000 – 2018 and 56.37% of 1950 – 2018 identified by systematic search).

Figure 2:

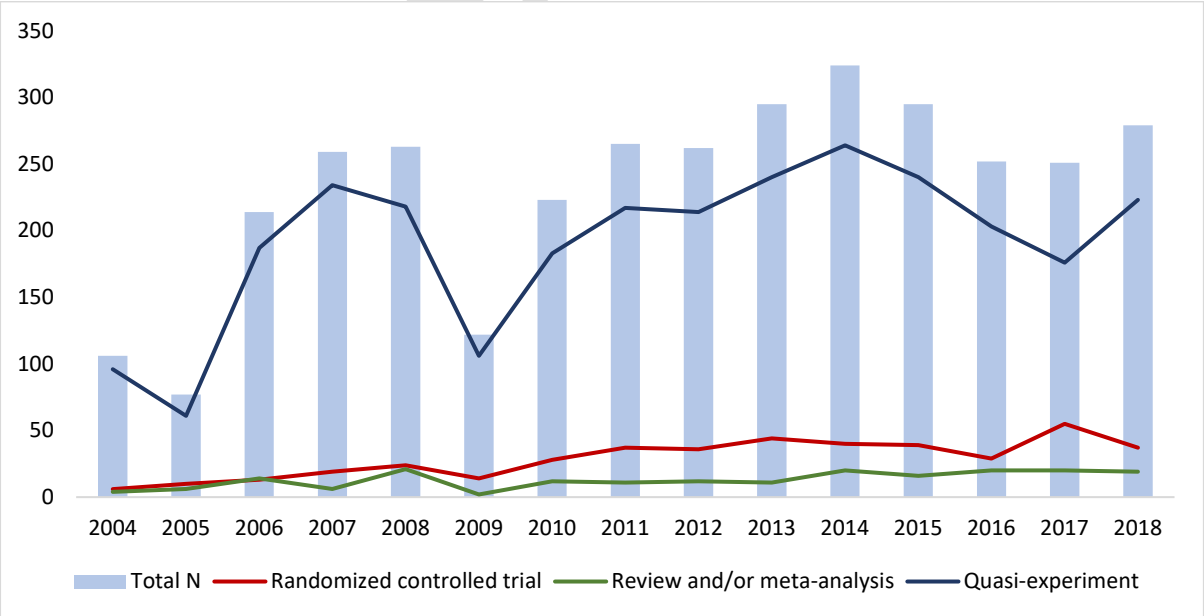
Types of Design*



**Note:* Includes data 2004-2018 only. N=3487

Figure 3:

Types of Designs over time*



**Note:* Includes data 2004-2018 only. N=3487

Table 1:
Types of interventions by study design for sample of eligible after Stage 2 screening based on reviews undertaken only (N = 1,149) [*]

Broad Intervention Category	**RCT**	**Review**	**Quasi-Experiment**	**Total**
Policing Practices	67	44	378	489
Investigative Techniques (Interrogation, Evidence Gathering)	60	6	92	158
Police Organisations, Staffing and Training	39	8	335	382
Police Technologies, Equipment & Safety	12	2	54	68
Legislative, Regulatory or Policy Reforms or Changes	3	3	46	52
Total by Evaluation Design*	**181** (16%)	**63** (5.5%)	**905** (78.5%)	**1149**

[*]*Note:* This includes a sample of all S2 eligible that are drawn from reviews and extractions we have conducted. The distribution of categories may change as more reviews are completed.

Table 2: ***Number of Studies and RCTs by Country****

Ranking	Country	Total N	RCT N
1.	United States	1822	216
2.	Multiple countries	248	13
3.	United Kingdom	241	43
4.	Australia	190	36
5.	Country not reported	158	41
6.	Canada	128	19
7.	Sweden	63	10
8.	The Netherlands	47	8
9.	India	42	5
10.	Norway	32	5
11.	Italy	29	2
12.	South Korea	28	0
13.	Turkey	26	1
14.	Mexico	24	0
15.	Brazil / China	23	1[5]
16.	Israel / Spain / Taiwan	22	4[6]
17.	Germany	20	5
18.	France / New Zealand / Nigeria	16	7[7]
19.	Slovenia	14	0
20.	Switzerland	13	1
21.	Japan	12	1
22.	Russia	11	0
23.	Belgium / South Africa	10	1[8]
24.	Finland	9	0
25.	Thailand	8	2
26.	Chile / Denmark / Ghana	7	2[9]
27.	Pakistan	6	0
28.	Argentina / Colombia / Czech Republic / Hong Kong / Iran / Portugal / United Arab Emirates	5	3[10]
29.	Malaysia / The Philippines / Sri Lanka	4	1[11]
30.	Greece / Indonesia / Iraq / Saudi Arabia / Tanzania / Trinidad & Tobago / Uganda / Ukraine	3	1[12]
31.	Afghanistan / Bolivia / Bosnia / Croatia / Estonia / Ireland / Poland / Singapore	2	2[13]

[5] This RCT was conducted in Brazil
[6] *N* = 3 were conducted in Israel, with the fourth from Spain
[7] *N* = 5 were conducted in France, with one from each of New Zealand and Nigeria
[8] This RCT was conducted in South Africa
[9] One RCT was from Ghana and the other from Denmark
[10] *N* = 2 were conducted in Portugal, with the third from Colombia
[11] This RCT was conducted in Malaysia
[12] This RCT was conducted in Uganda
[13] *N* = 1 conducted in Afghanistan and *N* = 1 conducted in Bolivia

32.	**Albania / Bahrain / Bangladesh / Bermuda / Brunei / Costa Rica / Cyprus / Ethiopia / French Ghana / Guatemala / Kazakhstan / Kosovo / Kuwait / Laos / Libya / Madagascar / Malawi / Malta / Nepal / North Korea / Qatar / Romania / Serbia / Sierra Leone / Slovakia / The Caribbean / Uruguay / Vietnam**	**1**	**1**[14]
Total		**3,487**	**431**

[*] ***Note:*** **Includes data 2004-2018 only. N=3,487**

[14] This RCT was conducted in Slovakia

References

Ater, I., Givati, Y., & Rigbi, O. (2014). Organizational structure, police activity and crime. *Journal of Public Economics, 115*, 62-71.

Bennett, S. & Newman, M. (2018). *Evidencing Police Training.* Paper presented at the ANZ Society of Evidence-Based Policing Conference, Australian Institute for Police Management (AIPM), Sydney, Australia.

Braga, A. A., & Bond, B. J. (2008). Policing crime and disorder hot spots: A randomized controlled trial. *Criminology, 46*(3), 577-607.

Braga, A. A., Turchan, B., Papachristos, A. V., Hureau, D. M. (2019). Hot spots policing of small geographic areas effects on crime. Campbell Systematic Reviews. doi: 0.1002/cl2.1046

Braga, A. A., Turchan, B., Papachristos, A. V., Hureau, D. M. (2007). Hot spots policing of small geographic areas effects on crime. Campbell Systematic Reviews. doi: 0.1002/cl2.1046

Braga, A. A., Welsh, B. C., Papachristos, A. V., Schnell, C., & Grossman, L. (2014). The growth of randomized experiments in policing: The vital few and the salience of mentoring. *Journal of Experimental Criminology*, *10*(1), 1-28.

Blattman, C., Green, D., Ortega, D., & Tobón, S. (2018). Hotspots interventions at scale: the direct and spillover effects of policing and city services on crime in Bogotá.

Collazos, D., García, E., Mejía, D., Ortega, D., & Tobón, S. (2020). Hot spots policing in a high-crime environment: An experimental evaluation in Medellin. Journal of Experimental Criminology, 1-34.

Eggins, L., Hine, L., Higginson, A. & Mazerolle, L. (2020). The impact of arrest and seizure on drug crime and harms: A systematic review. *Trends & issues in crime and criminal justice,* (602). Retrieved from https://www.aic.gov.au/publications/tandi/tandi602

Eggins, E., Hine, L., Mazerolle, L., McEwan, J., Hassall, G., Roetman, S., Roetman, S., & Bevis, K. (2020). *Mental health co-response models: a rapid review of the evaluation literature: final report* (for the New Zealand Evidence Based Policing Centre). Brisbane, Australia: The University of Queensland.

Eggins, E., Hine, L., McEwan, J., Hassall, G., Roetman, S., Roetman, S., & Mazerolle. L (2020). *Child safety inventions a rapid scoping review of the evaluation literature (Phase I & II)* (for the New Zealand Evidence Based Policing Centre). Brisbane, Australia: The University of Queensland.

Eggins, E., Mazerolle, L., Higginson, A., Hine, L., Walsh, K., Sydes, M., McEwan, J., Hassall, G., Roetman, S., Wallis, R., & Williams, J. (2019). *Criminal justice responses to child exploitation material offending: a systematic review and evidence and gap map* (for the Australian Institute of Criminology). Brisbane, Australia: The University of Queensland.

Eggins, E., Mazerolle, L., Hine, L., McEwan, J., Hassall, G., Roetman, S., & Roetman S. (2020). *Policing child sex offenders and offending - A rapid review of the evaluation literature: final report* (for the New Zealand Evidence Based Policing Centre). Brisbane, Australia: The University of Queensland.

Farrington, D. P., & Petrosino, A. (2001). The Campbell collaboration crime and justice group. The Annals of the American Academy of Political and Social Science, 578(1), 35-49.

Garner, J. H., & Visher, C. A. (2003). The production of criminological experiments. *Evaluation Review*, *27*(3), 316-335.

Gill, C., Weisburd, D., Telep, C., Vitter, Z., & Bennett, T. (2017). PROTOCOL: Community-oriented policing to reduce crime, disorder, and fear and improve legitimacy and satisfaction with police: a systematic review. *Campbell Systematic Reviews, 13*(1), 1-30.

Head, B. W. (2008). Three lenses of evidence-based policy. *Australian Journal of Public Administration, 67*(1), 1-11.

Heide, S., Stiller, D., Lessig, R., Lautenschläger, C., Birkholz, M., & Früchtnicht, W. (2012). Medical examination of fitness for police custody in two large German towns. *International journal of legal medicine, 126*(1), 27-35.

Higginson, A., Eggins, E., & Mazerolle, L. (2017). Police techniques for investigating serious violent crime: A systematic review. *Trends & issues in crime and criminal justice, 539*, 1–13. Retrieved from https://aic.gov.au/publications/tandi/tandi539

Hinkle, J. C., Weisburd, D., Telep, C. W., & Petersen, K. (2020). Problem-oriented policing for reducing crime and disorder: An updated systematic review and meta-analysis. *Campbell Systematic Reviews, 16*(2), e1089.

Kelling, G. L., Pate, T., Dieckman, D., & Brown, C. E. (1974). The Kansas City preventive patrol experiment. *Washington, DC: Police Foundation*.

Kelling, G. L., Pate, A., Ferrara, A., Utne, M., & Brown, C. E. (1981). The Newark foot patrol experiment. Washington. *DC: The Police Foundation*.

Kelling, G. L., & Wilson, J. Q. (1982). Broken windows. Atlantic monthly, 249(3), 29-38. Retrieved from https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/

Laycock, G. (2012). In support of evidence-based approaches: a response to Lum and Kennedy. *Policing: A Journal of Policy and Practice, 6*(4), 324-326.

Lum, C., & Kennedy, L. W. (2012). In support of evidence-based approaches: a rebuttal to Gloria Laycock. *Policing: A Journal of Policy and Practice, 6*(4), 317-323.

Lum, C., Kennedy, L. W., & Sherley, A. (2006). Are counter-terrorism strategies effective? The results of the Campbell systematic review on counter-terrorism evaluation research. *Journal of experimental criminology, 2*(4), 489-516.

Lum, C. M., & Koper, C. S. (2017). *Evidence-based policing: Translating research into practice*. Oxford University Press Oxford, UK.

Lum, C., Koper, C. S., Wilson, D. B., Stoltz, M., Goodier, M., Eggins, M., Higginson, A., & Mazerolle. L (2020). Body-worn cameras' effects on police officers and citizen behavior: A systematic review. *Campbell Systematic Reviews, 16*(3). doi:10.1002/cl2.1112

Mazerolle, L., Bennett, S., Antrobus, E., & Eggins, E. (2017). The coproduction of truancy control: Results from a randomized trial of a police–schools partnership program. *Journal of Research in Crime and Delinquency, 54*(6), 791–823. doi:10.1177/0022427817705167

Mazerolle, L., Bennett, S., Eggins, E., Higginson, A., & Antrobus, E. (2018). Legitimacy in policing: an updated Campbell Collaboration systematic review.

Mazerolle, L., Cherney, A., Eggins, E., Higginson, A., Hine, L., & Belton, E. (2020). PROTOCOL: Multiagency programmes with police as a partner for reducing radicalisation to violence. *Campbell Systematic Reviews, 16*(3). doi:10.1002/cl2.1110

Mazerolle, L., Eggins, E., Cherney, A., Hine, L., Higginson, A., & Belton. E (2020). Police programmes that seek to increase community connectedness for reducing violent extremism behaviour, attitudes and beliefs. *Campbell Systematic Reviews, 16*(3). doi.org/10.1002/cl2.1111

Mazerolle, L., Eggins, E., Bennett, S., Roetman, S., Hine, L., McEwan, J., Hassall, G., & Hockey, A. (2019a). Policing alcohol impaired driving. A rapid review of the evaluation literature.

Mazerolle, L., Eggins, E., Bennett, S., Roetman, S., Hine, L., McEwan, J., Hassall, G., & Hockey, A. (2019b). Road policing and level of centralisation: a rapid review of the evaluation literature.

Mazerolle, L., Eggins, E., & Higginson, A. (2016). PROTOCOL: Third party policing for reducing crime and disorder: A systematic review. *Campbell Systematic Reviews, 12(1).* doi.org/10.1002/CL2.153

Mazerolle, L., Eggins, E., Higginson, A. (2020) Street-level drug law enforcement: An updated systematic review. *Trends & issues in crime and criminal justice,* (599). Retrieved from https://www.aic.gov.au/publications/tandi/tandi599

Mazerolle, L., Eggins, E., Sydes, M., & Hine. L (2019) Policing domestic and family violence: A review of the evaluation literature. *Police Science, 4*(1), 18 – 20. Retrieved from https://www.anzsebp.com/wp-content/uploads/2020/02/ANZSEBPWinter19-compressed.pdf

Mazerolle. L., Eggins, E., Sydes, M., Hine, L., McEwan, J., Norrie, G., & Somerville, A. (2018). *Criminal justice responses to domestic and family violence: a rapid review of the evaluation literature: final report* (for the Queensland Domestic Violence Death Review Board). Brisbane, Australia: The University of Queensland.

National Academies of Sciences, Engineering, and Medicine. (2018). *Proactive policing: Effects on crime and communities*. National Academies Press.

Neyroud, P. W. (2017). *Learning to Field Test in Policing: Using an analysis of completed randomised controlled trials involving the police to develop a grounded theory on the factors contributing to high levels of treatment integrity in Police Field Experiments.* (Doctoral thesis). Retrieved from https://doi.org/10.17863/CAM.14377

Pierce, G. L., Spaar, S., & Briggs, L. R. (1988). The character of police work: Strategic and tactical implications. Center for Applied Social Research, Northeastern Unviersity.

Sherman, L. W. (1998). *Evidence-based policing*. Police Foundation Washington, DC.

Sherman, L. W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. Criminology, 27(1), 27-56.

Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P., & Bushway, S. (1997). Preventing crime: What works, what doesn't, what's promising: A report to the United States Congress. Washington, DC: National Institute of Justice.

Sherman, L. W., & Weisburd, D. (1995) General deterrent effects of police patrol in crime "hot spots": A randomized, controlled trial. *Justice Quarterly, 12*(4), 625-648. doi: 10.1080/07418829500096221

Sherman, L., Williams, S., Ariel, B., Strang, L.R., Wain, N., Slothower, M., & Norton, A. (2014). TTPS hotspot experiment: Murder, wounding, shooting data graphs. Unpublished manuscript. Retrieved from https://www.enterprise.cam.ac.uk/case-studies/hot-spot-policing

Sousa, W., Ready, J., & Ault, M. (2010). The impact of TASERs on police use-of-force decisions: Findings from a randomized field-training experiment. *Journal of experimental criminology, 6*(1), 35-55.

Telep, C. W., Garner, J. H., & Visher, C. A. (2015). The production of criminological experiments revisited: The nature and extent of federal support for experimental designs, 2001–2013. *Journal of Experimental Criminology*, *11*(4), 541-563.

Wilson, D.B., Mazerolle, L., & Neyroud, P. (2020). Campbell Collaboration systematic reviews and the Journal of Experimental Criminology: Reflections on the last 20 years. *Journal of Experimental Criminology.* doi:10.1007/s11292-020-09433y

Woody, W. D., Forrest, K. D., & Yendra, S. (2014). Comparing the effects of explicit and implicit false-evidence ploys on mock jurors' verdicts, sentencing recommendations, and perceptions of police interrogation. *Psychology, crime & law, 20*(6), 603-617.