

THE COST-EFFECTIVENESS OF EVIDENCE-BASED POLICING

Lawrence W. Sherman
University of Cambridge

One of the least-examined issues in government budgets is what police should do. Democratic governments across the world engage in endless debates about the cost of policing overall, but not about how the money is spent. The primary debate is always about how many police officers should be hired. More recently, legislative debates have examined how to manage the overtime budget.¹ These debates are generally premised on the theory of less police, more crime. Yet both debates overlook a far more important question, for taxpayers' money and for public safety: how to allocate police *time* to the most cost-effective *tasks*. The answer to this question can shift the very premises of these debates.

The purpose of this essay is to show how finance ministries and cabinets in parliamentary democracies can reduce *both* crime and police costs, simultaneously. By using a new strategy called "Evidence-Based-Policing," governments can work transparently with police and the public to invest in police tasks that are cost-effective. These investments can be made while cutting police budgets by putting an end to expensive but ineffective police tasks.

While this formula may sound too easy in principle, it is certainly challenging to accomplish in practice. What it requires for success is an open public deliberation about what works in policing, what doesn't, and how we know—the statistical research *evidence* that is the basis for making these difficult decisions. Rather than shrouding police work in the romance and mystery of detective fiction, governments must lead the public into the practical world of budgets, tasks, and results, generally expressed in the un-romantic form of numbers.² They must help the public answer the question of how police should divide up their personnel and their activities, within the limits of a *finite* amount of taxpayers' money.

Police leaders will often repulse anyone who raises this question by citing the many obstacles to answering it in a budgetary framework. Policing is too unpredictable, it is said. There are too many random combinations of urgent tasks to impose any rigid framework on a police officer's day. The importance of reserve capacity, like firefighters sleeping in the fire station, is too great for making police too busy with non-emergency

¹ England and Wales, for example, recently proposed to reduce the overall police budget by 2.5% by eliminating police overtime. See Home Office White Paper on Policing, December, 2009. This idea may be appealing in the abstract, but any suspension of a rape or murder investigation in full flight just because of an overtime ban could prove to be politically disastrous.

² Skillful politicians, however, might quote Nineteenth Century social reformers such as Florence Nightingale, the founder of modern nursing, who claimed that statistics were messages from God: "To understand God's thoughts we must study statistics, for these are the measure of His purpose". See <http://www.math.yorku.ca/SCS/Gallery/flo.html>

response. Even the setting of priorities among police tasks is said to be unrealistic, given the art of deciding minute-by-minute what needs to be done next—in a never-ending triage.

These claims are all true, up to a point. An English Chief Constable, for example, recently provoked press comment when he released a set of priorities. The news in the headline was this: police place alcohol as a priority higher than terrorism. The police force in question had no recent record of terrorist incidents, or even of generating local residents who committed terrorism elsewhere. Like most police agencies, however, it has a chronic problem of alcohol-fueled violence and road accident injuries. So if the general police goal was to reduce harm in the community, the Police Chief's priority was absolutely right—and the media criticism was entirely wrong. Moreover, if a terrorist incident actually did occur, we could predict with confidence that the police would re-assign constables away from the pub-fights and on to the trail of the terrorists and their associates. Because assignments can always be changed on very short notice, it still makes sense to set priorities for strategic allocation of police resources. Most police work can then be seen for what it is: what to do until the next emergency comes.

The point at which the objections become unreasonable is the argument that no strategy is possible except emergency response. Democratic police can well be trusted to take actions in emergencies; their record of doing so is nearly perfect, with a very few exceptions.³ What is less worthy of confidence is the police record of strategic allocations of personnel on a day to day basis. Here we have news media and politicians to blame for punishing police if they ever arrive late to a small but tragic emergency. That, in turn, makes police resist innovations that require sustained police engagement in activities that could systematically prevent crime, rather than just responding to it case-by-case.

There is a widespread public demand that the police be perfectly available at all times to arrive in 2 minutes at all locations. There is also a widespread demand for sunshine and pleasant weather 365 days a year, but that does not mean it is possible to meet that demand. People who expect such weather would be thought odd, unless everyone around them is equally unrealistic. Much the same can be said for expectations of policing. In both examples, the Hans Christian Anderson story of the Emperor's new clothes is reversed. In the case of unrealistic expectations, it is the Emperor who fails to point out that the *people* have no clothes. Until the government engages in openly democratic deliberation about what police do and what is truly possible or useful, police will be vulnerable to recurring attacks over their alleged “incompetence”—even though it is the critics who have been left incompetent to judge police performance by a lack of public discussion.

This essay has three parts. Part I describes *how* we know what is cost-effective in policing, against a wide range of goals and missions. It emphasizes the strong similarities between policing and public health, at least in the kinds of research methods used to

³ The summer, 1977 electricity blackout riots in New York City, where police failed to mobilize quickly, and the Rodney King trial verdict riots in Los Angeles in 1992, where police failed to take action to protect people and businesses are two notable exceptions. See the report of the Los Angeles Commission of Inquiry chaired by Warren Christopher, who subsequently served as the foreign minister of the US (“Secretary of State”) under President Clinton. The Commission's report led to a revolution of governance of the Los Angeles Police, with far greater accountability to the City's legislature.

generate knowledge for each field. Part II summarizes some of the key conclusions for *what* we know about cost-effective policing, first examining what does work, then examining what does not—at least based on evidence generated primarily in the US, the UK and Australia. Many other countries rely on the evidence from these countries in public safety as in public health, but can only be sure it applies in their own country by repeating some of the key research. Part III describes how any country can set up its own program for generating police knowledge, even with only a modest expenditure of research funding.

I. How We Know What Is Cost-Effective

The basic method by which we know what police tasks are cost-effective is to make fair comparisons between doing and not doing a police task. Those comparisons have three key elements. One is the true cost of a police activity, which can often be difficult to estimate. A second is the clear distinction between those places, times, or situations in which the task is performed and those of like character in which it is not: the “like-for-like” comparisons in which the only difference should be whether or not the police task in question is performed. The third is the definition of the goal or purpose of performing the police task: the intended outcome of police activity. This element is often the most difficult to specify, since many of the goals of policing may be achieved only by the cumulative impact of many police tasks over a long period of time. Nonetheless, broad consensus exists that police should accomplish certain measureable goals, such as preventing crime in high crime areas, identifying and prosecuting criminals, and reducing repeat offending among criminals who are convicted.

These three basic elements can produce what is called the *absolute* cost-effectiveness of a police task. The best example of this is the evidence from police strikes, such as the Helsinki study of 1976 (Makinen and Takala, 1980). In this study, observers were counting the number of fights they saw in late evening hours on the streets of Helsinki during a cold winter when the police went on strike, then came back to work. The counts of fights per night could then be turned into an average level for three distinct periods: 1) when normal levels of police patrol were provided; 2) when police patrol was vastly reduced to that of a few supervisors; 3) when police patrol was restored to pre-strike levels after the strike was over. In this research design, the like-for-like comparison was between having a police force and not. The third period was a useful check on whether the lack of policing during the strike altered the effect of police patrols (and emergency response) on street violence. As expected, the average number of fights observed increased sharply during the strike, and then declined to pre-strike levels after the strike was over. The study could even provide an estimate of the cost of policing per fight prevented, if only by comparing the cost of full police staffing in the area (around the central railroad station) in which the observations were conducted to the difference between the nightly averages of fights with and without full staffing.

The Helsinki police strike research is a useful way to make several points about how we know what works in policing. The most basic point is to refute the claim often made that “you can’t prove a negative,” including how many crimes did *not* happen because of policing. It is true that we cannot prove that something is impossible—as in the famous example of black swans, never found in Europe but eventually discovered in

Australia in the 18th Century. But that is logically very different from demonstrating that certain events were caused *not* to happen because the government spent money to achieve that goal. In order to demonstrate how much crime police caused not to happen, we need only subtract the average number of crimes during a period with police from the average during a period with police. It is this kind of measureable difference that is that provides *evidence* of crime prevention, if not total *proof* that an absence of so many crimes was caused by policing.

The second point is that there are better and worse ways of trying to prove that police cause a measureable difference in crime. The method used to estimate the benefits of policing in Helsinki had several advantages, but it was far from perfect. While the research “controlled” for differences in the weather by having a constant low temperature in all three time periods, researchers could not eliminate all other possible explanations for the increase in fights during the police strike. It is at least possible, for example, that just by coincidence a large number of sailors from Russia arrived in Helsinki just as the police strike began, and sailed away from Helsinki on the day the strike ended. Unless the researchers knew how to check on every such possibility, they could not even begin to report the data needed to “control” for such “alternative rival explanations.” For that reason, other research designs have been devised that allow researchers to eliminate rival explanations they never even thought of. Evidence-based policing is far better off with those research designs that eliminate unknown explanations than with research that leaves those rival explanations un-“controlled.”

The third point is that a “controlled” trial must have more than a single case, such as the withdrawal of police patrol in only one winter in Helsinki. In order to control for such unknown explanations as violent sailors arriving and leaving, a well-controlled study can repeat the comparison across hundreds of cases. If the same result is obtained almost all of the time, it is logically far less likely to be due to such coincidences. The best way to design such a repeated test is to identify a large sample of cases that are as similar as possible, and then to allocate a police task to around half of such cases but not to the other half. This allocation is best done with *equal probability* that each case will either get the police task, or not, by using a computerized random numbers formula.

The term used to describe this random assignment method is unfortunately misleading to both police and the public. The word “random” implies a haphazard or unsystematic approach to allocating the cases to have a police task or not. The truth is exactly opposite: an extremely systematic and rigid approach is required in order to give each case a fair chance to have either one condition of policing or another. If it were not for the unfortunate connotations of gambling, it would be better to call this a “lottery” method. But since that term suggests that researchers are gambling with people’s safety, researchers in many fields continue to call this method a “randomized controlled trial,” or “RCT”. Even that term is confusing in policing, since a trial is usually defined by police as the prosecution of a case in court. But such confusion can be quickly eliminated by defining terms clearly at the outset of any discussions of research “evidence”--another word that means something very different in science (as in facts that test a hypothesis) than in policing (as in facts that support a criminal charge against someone as guilty).

The fourth point is that a complete withdrawal of a police task is not necessary to provide evidence of police cost-effectiveness. While the *absolute* benefit of policing can only be tested by all-or-nothing comparisons, most police budgets are allocated on the

basis of *relative* benefits. We rarely ask whether police should be abolished. But we often ask whether police should be assigned to spend time in schools, in searching for drinking drivers, or in arresting people for domestic violence. Choosing how best to allocate police time can be based on comparisons of the crime prevention “yield” of different police activities. While there will always be political demands that police do “something” about a wide range of rips in the social fabric, the available tools of policing are far more effective at dealing with some kinds of problems than with others. Managing police resources to reduce costs and crime requires knowledge about that effectiveness, as well as political and democratic discussions of what that knowledge implies.

It is possible, for example, for police chiefs to make such statements as “we can prevent more crime per police officer by assigning police to do this task rather than that task.” There are two possible ways to generate the evidence needed to support such claims. One way is to use random assignment to make direct comparisons between two different ways police can deal with a recurring problem. In both cases, police are investing *some* (costly) time in a problem, but one method may often require less cost than another. It is usually more expensive, for example, to arrest and prosecute someone who has engaged in a barfight than to drive that person home to a family member who will promise to keep the fighter from leaving home again that night. If both methods have a similar outcome in terms of subsequent violence and injury, police may be justly claim that the informal resolution is more cost-effective *at preventing harm*.

Harm reduction may not be adequate evidence, however, to support informal actions before an elected government committed to counting more “offenses brought to justice,” as the British government has been for the first decade of the 21st Century. If the goal is retribution, then the research can simply change its outcome measure. If criminal convictions in court is the measure of an offense brought to justice, the question then becomes whether one police task can produce convictions more cost-effectively than another. The use of DNA evidence in solving burglaries, for example, was the subject of a randomized controlled trial (RCT) recently conducted in the US (Roman, et al, 2008). In this RCT, the presence (and cost) of a trained evidence technician was randomly assigned to about half of a large sample of burglary investigations. The rate at which burglary suspects were identified and prosecuted was then compared between burglary cases with and without the assignment of evidence technicians.

Note that the DNA collection RCT measured both *absolute* and *relative* benefits of the cost of the technicians. Both kinds of benefit were based on the same retributive criterion of offenses brought to justice (and not crimes prevented). The *absolute* benefit was estimated by comparing *some* DNA collection to *none*. Financial analysts could then compare the *relative* benefit of adding the cost of DNA collection to the existing cost investigating each burglary, of which DNA collection was only a small additional fraction.

The two kinds of knowledge about police cost-effectiveness are therefore complementary. The absolute comparison of doing something to doing nothing can be useful to know, but is often not wise for police to do nothing about a problem. It is certainly unwise to withdraw police entirely from a city just to estimate their cost-effectiveness, given what we already know from a series of police strikes (Sherman, 1992a). But it is even more unwise to just “do something” without comparing different options of what to do, especially if some options can actually make things worse. While

we often learn best what does not work by making absolute comparisons, we can also compare relative benefits of two ways of police dealing with a problem.

The “Scared Straight” program, for example, has been tested largely by absolute comparisons. These comparisons showed that the program is not only a waste of money, it is also a cause of crime. This program, which sometimes involves police, takes early adolescents to a prison where they meet with “scary” criminals who tell them how dangerous it is in prison. This “logical” approach to doing “something” about juvenile delinquency has now been subjected to repeated RCTs (Petrosino, et al, 2003). The net effect of these RCTs shows an *increase*, rather than decrease, caused by the Scared Straight program in the rate at which the young people go on to be charged with crimes.

Discovering such important results is also made possible by RCTs that make relative comparisons. Comparing two different ways in which police respond to domestic violence, for example, requires that police always do something about an incident of domestic violence (Sherman and Berk, 1984; Sherman, 1992b). But RCTs can compare two very different ways of responding. One way can involve mediation or asking the offender to leave the home for the night. Another way is arrest the suspect and take him (or her) to jail for the night. The relative *benefit* of these two methods can be calculated by comparing the average repeat offending rates in large groups of cases in which either arrest or mediation was the police response. The relative *cost-effectiveness* of the two methods can then be compared on the basis of the average cost of police time invested in each case under the two methods. These computations are radically different if the outcome criterion is altered from repeat incidents to convictions in court. Yet it is important to know that a policy of arrest for domestic violence, just like a policy of “Scared Straight,” can cause more future crime than other police actions (Sherman 1992b). This fact adds even more cost (including medical treatment) to that method beyond the price of police time.

The fifth and final point on how we know what works is the idea of repeating randomized controlled trials, so that the average effect of a police method can be estimated across a wide range of samples. This method is called a *meta-analysis*, or a study of studies, to see what can be concluded by integrating their results. The best meta-analysis is based on a *systematic review* of *all* the evidence that meets a certain standard of scientific methods, and not just a selection of studies (which can be biased in favour of one or another result). As the Norwegian-based, international Campbell Collaboration has demonstrated, we can obtain very different results from systematic and unsystematic reviews of research (see www.campbellcollaboration.org). That is why even unpublished research can be vital. If studies that find “no effect” are left unpublished because they were “disappointing,” the average effect of a police practice found in only the published results could be a poor basis for predicting the cost-effectiveness that any police agency will obtain from that practice. Such “publication bias” often over-estimates the benefit of the practice, even with good research designs. Even with equally good scientific methods in the published and unpublished research, the selective publication of more positive results can clearly mislead police leaders and governmental financial analysts.

This would not be the case, however, with studies omitted from a systematic review based on criteria about the scientific methods each study uses. These criteria may vary, but most begin with the question of whether the research creates a fair comparison, or “unbiased estimate.” Fair comparisons have what is called “internal validity,” or a high

degree of similarity between cases treated in one way and cases treated in another (Campbell and Stanley, 1963). This baseline similarity helps reduce the range of alternative explanations of any apparent benefit from using one practice rather than another.

A widely used scale of rating internal validity of evaluation evidence was developed for the US Congress (Sherman, 1997a), and has been adopted in the UK and other countries. The scale ranks internal validity of comparative effectiveness studies from 1 to 5, with a “5” indicating the highest level of internal validity (produced by a well-implemented RCT). A “4” indicates a relatively large sample of “matched” (but not randomly assigned) cases that differ only in terms of the policies applied to them. A “Level 3” study is one that compares one unit (such as central Helsinki) before and after a policy change, but with a further comparison to at least one other similar unit (like central Stockholm) that had no such change in the same time period. A “Level 2” study is one that lacks any comparison to a similar unit; the Helsinki study (Makinen and Takala, 1980) itself was a Level 2. The weakest form of evidence about causation is a Level 1 study, which is simply a correlation between crime rates (or other criteria) with differing practices or policies.

In summary, *how* we know about what is cost-effective in policing is by making comparisons. The more internally valid we make these comparisons, the more confidence we can have that it is the choice of police actions, and not some other factor, that is causing any apparent benefit. If we can generate repeated tests that make the same comparison, we can then achieve what is called “external validity:” the capacity to generalize from results of existing research on certain kinds of populations and organizations to others in which no tests have yet been conducted.

While we know a great deal about internal validity, we must admit that far less is known about external validity (or “generalizability”). Whether knowledge from tests of policing in US cities can be generalized to Swedish or British cities is still unknown. The only way to determine the external validity of police research is to repeat the same test in different countries. While it is less expensive merely to apply the results of US studies, it may be far less cost-effective than repeating key studies in each country. In the long run, democracy and public safety will be better served by having as many nations as possible contribute to a common pool of knowledge of the cost-effectiveness of police actions with common patterns of crime.

II. What We Know About Effective Policing

What we know about effective policing has been the subject of repeated reviews since the late 1990s. Most of these reviews have been surveys of published findings across a range of questions (e.g., Sherman, 1997; Skogan and Frydl, 2004; Sherman and Eck, 2005). More recently, a number of systematic reviews of repeated tests of specific police practices have been published, both in scientific journals and on the website of the Campbell Collaboration. Some of the reviews are limited to Level 5 evidence (RCTs), while others accept evidence as low as Level 2. The summary of this evidence below is more illustrative than exhaustive, and varies in the level of both internal and external validity. All of the evidence reported here is arguably better than having no evidence at all, even if it is later contradicted by better or more particularized evidence. Any initial

finding about the effect of a general tool, such as the impact of shoes (versus bare feet) on health, for example, is likely to be followed by the later discovery that one size does not fit all. The vision of growing evidence is one that refines and particularizes knowledge to suit specific settings or circumstances, thus increasing the cost effectiveness of investing in both evidence and the practices the evidence indicates.

In one sense, this research glass is almost empty, in relation to all the facts we would like to know about the effectiveness of different police actions. Yet in another sense the glass is surprisingly full, considering just how recently any governments have invested in generating evidence about police outcomes. There is ample material available to be applied to policing, much of which is not yet taught in police academies or put to work in finance ministries. If the police profession will begin to use this evidence to make policing more cost-effective, that could create a tipping point (Gladwell, 2000) for investing in more evidence that will steadily lead to even greater cost-effectiveness.

So what do we know?

Patrol. We know that there are substantial benefits of focusing police patrols in high-crime hot spots (Sherman and Weisburd, 1995). Because some 3% to 5% of small places (buildings or street-blocks) in most US cities produce over half of all crime (Sherman, et al, 1989; Weisburd, et al, 2004), the benefits of concentrating police patrols in those areas seem theoretically powerful. In repeated RCTs, that theory has stood up, as Braga's (2007) independent systematic review of the evidence for the Campbell Collaboration has concluded. What his review does not yield is 1) a precise estimate of the cost-effectiveness of adding patrol time to crime hot spots, and 2) a review of the evidence about displacement effects from concentrated hot spot patrols. But the evidence is useful despite these limitations. Other evidence (Koper, 1995) even suggests the optimal benefit for patrol in a hot spot is obtained after only 15 minutes, at which point a patrol officer can move to another hot spot with a "residue" of crime prevention left behind. Finance ministry economists in any nation can use the US studies to estimate cost-effectiveness of crime prevention based on local police costs. And the issue of displacement has been addressed separately, in work by David Weisburd and his colleagues that won the Stockholm Prize in Criminology for 2010.

Displacement. In a series of tests associated with analysis of crime hot spots, Weisburd and his colleagues (2004, 2006) have shown that concentrating patrols in hot spots does not simply "push crime around the corner." In a study of crime over 15 years in Seattle, Weisburd's team (2004) reported that almost all the overall decline in crime during that period occurred in the blocks that had the greatest concentrations of crime. The 5% of blocks that had 50% of crime at the outset showed substantial reductions in crime, without increasing crime in the other 95% of blocks that remained stable with low crime rates. This was suggestive of crime reductions without displacement, with the longest time series of block-specific data subjected to that kind of analysis. What was even more powerful, however, was evidence from Jersey City (Weisburd, et al, 2006) showing steep reductions in drug dealing and prostitution without any crime increases in the area immediately surrounding the target areas. Instead of "displacing" crime to those buffer areas, the evidence actually showed a reduction in crime there as well. Weisburd calls

this a “diffusion of benefit,” in which the apparent deterrent effect of policing at the epicentre of a hot area spills over into surrounding areas as well.

Domestic Violence. Police responses to domestic violence are a prime exhibit for the case that one size does not fit all. In an initial RCT in Minneapolis, Sherman and Berk (1984) reported that on average, arresting suspects for domestic violence yielded less repeat violence over six months than either of two informal alternatives to arrest. When those authors succeeded in persuading the US Department of Justice to fund five further studies, however, the results became more complex. Sherman (1992b) showed that in three other cities, the effect of arrest depended on whether the suspect was employed. If the suspect had a job, according to the victim, the rate of official repeat calls (generally by other people besides the victim) was cut in half by an arrest compared to a warning. If the suspect was unemployed at the time of the incident, however, making an arrest doubled the measured rate of repeat offending. This finding was initially discovered in Milwaukee, in a primarily African-American population in a concentrated poverty area. It was then confirmed in a primarily white population in Omaha, and again in a primarily Hispanic community in the Miami area. Thus the evidence for differential effects of policing for employed and unemployed offenders—that one size does not fit all—held constant regardless of race or ethnicity. This evidence was not well received politically, in part because no one wished to confront its complex implications. But a further series of three Cambridge University RCTs under way in English police agencies in 2010 will soon extend the evidence to compare mandatory arrest to victim consultation as a basis for police responses to similar kinds of cases.

Gun Violence. In the high-homicide rate nations of the US and Colombia, consistent evidence has emerged that police can reduce gun violence by stop-and-search procedures. In their review of the mostly Level 3 studies, Koper and Mayo-Wilson (2006) concluded that there is some basis for predicting that stopping people in gun crime hot spots can reduce gun injury or homicide. This police strategy can be very controversial, depending upon how it is done. But the evidence from systematic tests reveals no rebellions provoked by police when focused on gun crime per se. What is most unclear about these studies is how they might generalize to countries where very few people carry or own guns, and whether stop-and-frisk patrols would be cost-effective without identifiable areas of concentrated gun crime.

Juvenile Crime. A long series of Level 5 studies (as well as less rigorous evidence) supports the conclusion that prosecuting most juveniles causes more crime rather than less. In a recent systematic review, Petrosino et al (2010) report that the average effect of prosecuting juveniles formally in court is to increase their offending rates, as compared to diverting them to informal dispositions that leave no criminal record. These studies are unusually cross-national, with RCTS on three continents contributing to the meta-analysis (US, UK and Australia). While other evidence suggests that very young persons charged with violence are at very high risk of later committing a homicide (Berk, et al, 2009)—at least in a city with a high rate of gun crime (Philadelphia), it is also true that very early onset of crime predicts both higher seriousness and frequency of offending through age 45 (Farrington, et al, 2006), even in a city (London) with a low rate of gun

crime. The evidence-based policing response to this evidence is challenging, but ready to be applied. Can police explain to the public why most juvenile cases should not be prosecuted? If so, can they organize a cost-effective strategy for diverting juvenile offenders (Sherman et al, 2000; Shapland et al, 2008) to restorative policing?

Restorative Justice. One of the most comprehensively tested and cost-effective policing strategies is police-led, face-to-face meetings between admitted offenders and crime victims, with their respective families and friends in the room. These two-hour meetings have been tested in Australia, the US and England. With both juveniles and adults, they have succeeded in reducing repeat offending—especially in the UK, where the police created specialized units to deliver restorative justice conferencing (RJC) rather than allowing field officers to handle cases, as the Australian Federal Police did. Almost 3,000 cases have been placed into 12 RCTs around the world (Strang and Sherman, 2009), with 10 out of 12 tests reducing repeat convictions over a two-year follow-up. While the standardized mean difference is relatively modest, the cost-effectiveness is very large. In London, for example, Shapland and her colleagues (2008) estimated that the return on investment was almost 14 to 1. Comparing the costs of crime prevented by RJC to the costs of delivering it, the Shapland team reported that across seven RCTs with different kinds of offenders and offenses, RJC produced an average return on investment of 8 to 1. Despite these impressive findings, however, it must be noted that police agencies have not yet applied this evidence to practice.

Public Drug Dealing. The problem of open air and private house drug markets that swept the US in the late 1980s provoked a series of RCTs. From New Jersey (Weisburd and Green, 1995) to California (Mazerolle, et al, 2000; Eck and Wartell, 1998), these studies found that police could effectively suppress public drug dealing with concentrated patrols and with building code enforcement. Less successful was the dramatic “raid” strategy in Kansas City, where police battered down the doors of crack houses with a battering ram. The RCT testing this strategy (Sherman and Rogan, 1995) found that the effect was big but short-lived, dramatically reducing crime on the block for only 12 days. More sustained control by police presence is more successful, but not necessarily more cost-effective (Sherman, 1990). Few economists have added the cost analysis to the available impact estimates.

Repeat Offenders. Another effective but expensive strategy is covert police surveillance of high-rate repeat offenders, many of whom have only recently returned from prison. The Washington DC RCT of this strategy showed that 24-hour surveillance of repeat felons targeted by police intelligence increased their odds of being arrested by some 400%, in comparison to similar offenders who were not put under surveillance (Martin and Sherman, 1986). The difference was almost entirely due to the surveillance team observing the offender committing a crime and arresting him *in flagrante delicto*. The evidence for the crime was then given by multiple police witnesses, leading to a high rate of conviction and re-incarceration. The costs of imprisonment are also a part of this analysis. Depending on the estimate cost and frequency of the crimes the offenders could have committed outside of prison, the strategy may or may not have been cost-effective. Future studies should place more emphasis on the total cost of crime committed in the

control group as well as the targeted group, in order to refine the cost-effectiveness estimates.

Investigations. In addition to the DNA collection RCT (Roman, et al, 2008) cited above, several other studies of investigations have been undertaken. Perhaps the most useful is the research on “solvability scores,” by which an investigator can decide in the first few minutes of considering a case whether it is likely to result in an arrest. Using thousands of cases to identify key predictors of likely success—such as whether an eyewitness has been identified—an investigator can perform a kind of triage to invest her time most usefully in bringing offenders to justice. Investing time in cases that are unlikely to produce an arrest no matter how much time investigators take is arguably a waste of money. But it is also arguably a fact that should be discussed with government and the public. The idea of a “right” to have every case investigated, no matter how hopeless, is very expensive, draining money from more productive efforts to create a safe society. Making such decisions is what democracies can do better when they are informed by better evidence.

Knowledge Gaps. What we don’t know about police effectiveness still outweighs what we do know. This includes both entire subjects that have never been studied (such as domestic terrorism prevention) and particular issues in topics that have already been studied (such as how women respond to arrest for domestic violence against men). One of the most crucial remains the displacement of individuals, as distinct from criminal events, from areas of concentrated patrol to other areas without such patrols. Other questions include the value of police participation in joint efforts with probation and child welfare agencies, where “multi-agency partnerships” have become fashionable without any evidence. The cost of police attending meetings with other agencies remains demonstrably high. The basis for calculating a return on investment remains entirely lacking. The same is true for a vast array of expensive police activities, including all training, the management of dispatch centers, and time spent filling out paperwork.

Ultimately, the assessment of what we know, what we don’t, and what should be studied next is a question for police and the public. Academics may currently drive the choice of research questions, but only because no one else is pushing for such research. Filling the gaps in evidence-based policing is something that should begin with research consumers—the police and public—and then be handed over to skilled researchers. That is the basic idea by which any nation can develop its own evidence base for cost-effective policing.

III. How Any Country Can Develop Evidence-Based Policing

This chapter reflects my experience in police policy research over four decades on three continents. While no country has ever developed evidence-based policing in a way that approaches the model of evidence-based medicine or public health, there are distinct contributions that different countries have made. The US, for example, has funded and completed the majority of all experiments in policing. It has done little, however, to apply the results of those tests to police practice. The UK, in contrast, has conducted very few rigorous tests of police practice, but has done more to promote the virtues of evidence-

based decisions than police in any other nation. Depending on the management of impending budget cuts in policing, the UK may make substantial progress in applying evidence as a major criterion in choosing what to spend money on, and what to cut.

These reflections are informed by the major progress in 2008-9 at Cambridge University, where the combination of a postgraduate course for police with a new research centre has spawned a booming research partnership with British police agencies. Rather than describing the particulars of those developments, the following abstracts the key elements from the UK experience to offer a universal plan for any nation to follow.

Create university-based expertise in experimental criminology. Experimental criminology is a rapidly growing field, with its own journal (JOURNAL OF EXPERIMENTAL CRIMINOLOGY, founded in 2005), its own honorary academy (ACADEMY OF EXPERIMENTAL CRIMINOLOGY, founded in 1998) and its own division of the American society of Criminology (the Division of Experimental Criminology, founded in 2009). The number of social scientists conducting experiments world-wide is growing, and any country that wishes to can attract or develop its own expertise in this area. Once a university—or even a government research institute—has a capacity to design, deliver and analyze randomized field trials (as distinct from laboratory experiments, which are of questionable external validity to field settings), a nation can direct that capacity to support the development of local knowledge about police effectiveness. This can begin by the simple process of replicating experiments already conducted in other countries. It can then move on to designing new studies never done anywhere before.

All of these efforts can be supported by the Greater Manchester Police (GMP) partnership with Cambridge University's new Jerry Lee Centre of Experimental Criminology, which has posted protocol templates for designing experiments and provided a public registry for declaring an experiment to be underway—an important antidote to publication bias.⁴ This partnership, called the Tactical Experiments and Strategic Testing Program (TEST), has now attracted several other large police agencies in the UK, which will simultaneously conduct the same experiments in different sites across Britain. This will vastly expedite the process of testing for external validity, in order to know whether what works in one community will work in others as well.

2. Have universities share that knowledge with senior police leaders.

It seems essential for evidence-based policing that academics and police leaders speak the same language of evidence. One way to do this is for universities to offer post-graduate degrees for senior police leaders, in which the curriculum focuses heavily on the production and interpretation of research evidence on police effects. A master's degree on such knowledge can be offered on a part-time “executive” basis that allows senior police to earn a degree while continuing in their career. In the process of learning how to use these tools, it is entirely possible that some leaders will seize the initiative to fill gaps in police knowledge—driving the production of the evidence police need and will use.

⁴ All of this and other information can be found at www.crim.cam.ac.uk/experiments

3. Ask police leaders to choose topics and research questions.

By planting the seeds of curiosity among police leaders, universities can gain from having police choose the topics and research questions. Ministries of finance can also push for evidence on certain questions, such as those related to costs and choices. Once the police have a high sense of priority about doing research, complicated experiments become much more feasible to complete.

4. Allow universities to control the research design and data analysis.

With support from top police who want answers to their questions, academics can be empowered to require high quality research designs. They can be given full opportunities to explain why Level 5 is so much better than Level 3, and why an experiment must be designed in a particular way with a particular sample size. If academics are clearly given independence in saying what is needed for a valid study, more credence will be given to the results—just as in medical trials.

5. Have police and universities jointly report on their conclusions to government and the public.

Once an RCT or a series of RCTs has been concluded, the participating academics and police leaders can publish the results together, with police as co-authors. Incentives for senior police promotion can be tied to participation in (and co-authorship of) operational field effectiveness studies. The development of professorships in policing can further attract police leaders to do research, and to collaborate with fulltime academics. In the long run, research can become embedded as a central part of the police task. All of this can enhance the capacity of police to tell the public what is cost-effective for them to do—and what is not. At a higher level of professional consensus, police leaders who have not been a part of any specific research project can serve as assessors of the strengths and weaknesses of the research, as well as what policy conclusions might be drawn from it.

By this model, any country could rely on relatively modest investments in a few universities to draw on the existing budgets of police agencies to produce new knowledge. Most of the cost of police experiments lies in police salaries. Investing those salaries in learning more about crime prevention is something that many police agencies have done for years, from Minneapolis to London. If the idea of evidence can now spread around the police world as fast as it did for medicine generations ago, the safety of our democracies may improve as rapidly as our health has done.

REFERENCES

References

Richard Berk, Lawrence Sherman, Geoffrey Barnes, Ellen Kurtz and Lindsay Ahlman (2009). “Forecasting Murder within a Population of Probationers and Parolees: A High Stakes Application of Statistical Learning.” *Journal of the Royal Statistical Society. Series A. Statistics in Society* 172:191-211.

Braga, Anthony (2007). "The Effects of Hot spots Policing on Crime." A Campbell Collaboration Systematic Review. <http://www.campbellcollaboration.org/library.php>.

Campbell, Donald T. and Julian C. Stanley (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand-McNally.

Eck, John, and Julia Wartell (1998). Improving the management of rental properties with drug problems: A randomized experiment. *Crime Prevention Studies, Vol. 9*, (pp. 161-185). Monsey, NY: Criminal Justice Press.

Farrington, David, Jeremy W. Coid, Louise M. Harnett, Darrick Jolliffe, Nadine Soteriou, Richard E. Turner and Donald J. West (2006). Criminal careers up to age 50 and life success up to age 48: new findings from the Cambridge Study in Delinquent Development. 2nd edition. Home Office Research Study 299. London: Home Office.

Gladwell, Malcolm (2000). *The Tipping Point*. Boston: Little, Brown.

Koper, Christopher (1995). "Just Enough Police Presence: Reducing Crime and Disorderly Behavior By Optimizing Patrol Time In Crime Hot Spots." *Justice Quarterly* 12: 649-672.

Koper, Christopher and Evan Mayo-Wilson (2006). "Police crackdowns on illegal gun carrying: a systematic review of their impact on gun crime" *Journal of Experimental Criminology* 2: 227-261.

Makinen, Tuija and Hannu Takala (1980). "The 1976 Police Strike in Finland." *Scandinavian Studies in Criminology* 7:87-106.

Martin, Susan and Lawrence Sherman (1986). "Selective Apprehension: A Police Strategy for Repeat Offenders." *Criminology* 24: 55-72.

Mazerolle, Lorraine Green, James F. Price and Jan Roehl (2000). "Civil Remedies and Drug Control: A Randomized Field Trial in Oakland, California." *Evaluation Review* 24: 212-241.

Petrosino, Anthony, Carolyn Turpin-Petrosino, and John Buehler (2003). Scared straight and other juvenile awareness programs for preventing juvenile delinquency. (Campbell Collaboration Review). In: *The Campbell Collaboration Reviews of Intervention and Policy Evaluations (C2-RIPE)*, November, 2003. www.campbellcollaboration.org

Petrosino, Anthony, Carolyn Turpin-Petrosino, and Sarah Guckenberger (2010) "Formal System Processing of Juveniles: Effects on delinquency." Campbell Collaboration Review. <http://www.campbellcollaboration.org/library.php>.

Roman, John K., Shannon Reid, Jay Reid, Aaron Chalfin, William Adams and Carly Knight (2008). *The DNA Field Experiment: Cost-Effectiveness Analysis of the Use of DNA in the Investigation of High-Volume Crimes*. Washington, DC: Urban Institute.

Shapland, Joanna, Anne Atkinson, Helen Atkinson, James Dignan, Lucy Edwards, Jeremy Hibbert, Marie Howes, Jennifer Johnstone, Gwen Robinson and Angela Sorsby (2008). *Does restorative justice affect reconviction? The fourth report from the evaluation of three schemes*. London: Ministry of Justice.

Lawrence W. Sherman (1990). "Police Crackdowns: Initial and Residual Deterrence" in Michael Tonry and Norval Morris, eds., *Crime and Justice: an Annual Review of Research*, Volume 12, pp. 1-48. Chicago: University of Chicago Press.

Lawrence W. Sherman (1992a). 1992 "Attacking Crime: Police and Crime Control." in Norval Morris and Michael Tonry, eds., *Modern Policing: Crime and Justice*, Vol. 15, pp. 159-230. Chicago: University of Chicago Press.

Sherman, Lawrence (1992b). *Policing Domestic Violence: Experiments and Dilemmas*. NY: Free Press.

Lawrence W. Sherman and Dennis P. Rogan (1995). "Deterrent Effects of Police Raids on Crack Houses: A Randomized, Controlled Experiment" *Justice Quarterly*, Vol. 12, No. 4.

Sherman, Lawrence W. (1997a). "Thinking about Crime Prevention." Chapter 2 in Lawrence W. Sherman, Denise Gottfredson, Doris MacKenzie, John Eck, Peter Reuter and Shawn D. Bushway, *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, DC: US Department of Justice (<http://www.ncjrs.gov/works/>)

Sherman, Lawrence W. (1997b). "Policing for Crime Prevention." Chapter 8 in Lawrence W. Sherman, Denise Gottfredson, Doris MacKenzie, John Eck, Peter Reuter and Shawn D. Bushway, *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, DC: US Department of Justice (<http://www.ncjrs.gov/works/>)

Sherman, Lawrence (2010). *Experimental Criminology*. London: Sage.

Sherman, Lawrence and Richard Berk (1984). "The Specific Deterrent Effects of Arrest for Domestic Assault" *American Sociological Review*, 49(2): 261-271.

Sherman, Lawrence Gartin and Buerger (1989). "'Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place." *Criminology*, Vol. 27, No. 1, 27-55.

Sherman, Lawrence and Weisburd (1995). "General Deterrent Effects of Police Patrol in Crime Hot Spots: A Randomized, Controlled Trial." *Justice Quarterly*, Vol. 12, No. 4: 635-648.

Sherman, Lawrence W., Heather Strang and Daniel J. Woods (2000). *Recidivism Patterns in the Canberra Reintegrative Shaming Experiments (RISE)*. Canberra: Centre for Restorative Justice, Research School of Social Sciences, Australian National University.

Sherman, Lawrence W. and John Eck (2005). "Policing For Crime Prevention". In L.W. Sherman, David P. Farrington, Brandon Welsh and Doris MacKenzie, eds., *Evidence-Based Crime Prevention*. London: Routledge.

Skogan and Frdyl, eds. (2004). *Fairness and Effectiveness in Policing: The Evidence*. Washington, D.C.: National Academies Press.

Strang, Heather, and Lawrence Sherman (2009). *Effects of Restorative Justice Conferencing on Victims and Offenders: A Systematic Review*. Draft Review for the Campbell Collaboration. Cambridge University, Institute of Criminology, Jerry Lee Centre of Experimental Criminology.

Weisburd, David and Lorraine Green (1995). "Policing Drug Hot Spots: The Jersey City Drug Market Analysis Experiment." *Justice Quarterly* 12: 711-735.

Weisburd, David, Shawn Bushway; Cynthia Lum; Sue-Ming Yang (2004). "Trajectories of Crime at Places: A Longitudinal Study of Street Segments in the City of Seattle." *Criminology* 42: 283-322.

Weisburd, David, Laura Wyckoff, Justin Ready, John E. Eck, Joshua C. Hinkle, and Frank Gajewski (2006) "Does Crime Just Move Around the Corner?: A Controlled Study of Spatial Displacement and Diffusion of Crime Control Benefits." *Criminology* 44(3), 549-591.