Evidence-Based Crime Prevention: 
A Global View from the U.S. to Japan

Lawrence W. Sherman 
University of Pennsylvania

Abstract

How can democracies prevent crime without causing crime? How can they avoid wasting money on programs that sound good but have no effect? How can they make decisions based on rational analysis, rather than on emotional impulses? These problems afflict all democracies, regardless of culture or history. They may also have a common solution in evidence-based government. This article reviews the concept of evidence-based government and its DRIVER model, as well as the standards of evidence for different kinds of questions about crime prevention. Using these standards, it suggests that while randomized controlled trials are needed for only a small portion of the evidence in crime prevention, RCTs in each country are indispensable in order to keep programs from causing more crime than they prevent. The leadership of advanced nations is needed to move RCTs forward to a more rational framework for crime prevention policy.

Key words: evidence, experiments, crime prevention

Democracies tend to deal with crime prevention policy on a crisis-by-crisis basis. While this approach often produces substantial changes in crime prevention policy, there is great doubt that such policy changes prevent crime. The policy changes may increase the severity of punishment. The changes may increase the level of public spending designated for crime prevention. The changes may even persuade voters that the elected government is doing a good job. None of these changes after a crime crisis, of course, constitutes success in crime prevention. As Winston Churchill once said, “However beautiful the strategy, you should occasionally look at the results.”

The 2004 murder of Satomi Mitari, age 12, by an 11-year-old fellow student at Okubo Elementary School in Sasebo is a tragic example of such a crime that can cause such crisis (McAvoy, 2004). The 2003 murder of a 4-year-old in Nagasaki (by a 12-year-old boy) is another. Both are reminiscent of the murder of 2-year-old James Bulger near Liverpool by two 10-year-old boys in 1993, not to mention the 1999 murders of 12 students and a teacher by two students at Columbine High School in Colorado. None of these crimes may be clearly identified as the cause of a major shift in crime prevention policy. But all of them help to create a context of crisis for crime prevention policy. Elected officials must know that if the policy is not changed in
response to the present crisis, they will find it all the more difficult to defend their policies in government when the next crime crisis erupts like a social volcano.

Thus elected officials predictably react to crime crises in ways that appear emotional and moral, rather than analytic and rational. As the French sociologist Durkheim first observed in the 1890s, crime creates a moment for reinforcing social solidarity by condemning the behavior of those who symbolize the anti-thesis of the society’s values. The murder of a schoolchild is an outrageous act that is a clear demonstration of UN-Japanese behavior, as well as un-American and un-British. Elected officials resort first to such statements, as some modern scholars say they should (e.g., Braithwaite, 1989), as a means of re-affirming moral values. But then they may go on to seize the first idea that appears on their desk, and rush it into action. Such decisions go beyond what is emotionally necessary to help a grieving population; an equally satisfactory response could be one that is known to be effective, rather than one that appears to be effective.

Yet in the immediate aftermath of a crime crisis, politicians may not think they have time to wait for a thorough review of the cost-effectiveness of different alternatives for preventing crime. They may move quickly to increase prison space or the length of prison terms. They may change the law to prosecute children as adults. They may seek to increase police powers to investigate or detain crime suspects. They may even spend billions of Yen to pay for programs for young people, such as “Scared Straight” or “Drug Abuse Resistance Education,” because they look good on television, even though the research evidence on such programs show that they are either ineffective or may even cause crime. Most of all, they may resort to the “get tough” approach, on the grounds that the emotions of retribution provide public satisfaction for the reaffirmation of morality. But as one former U.S. elected official has observed, “Politics is not religion, and we should govern on the basis of evidence, not theology” (Clinton, 2004).

In August of 2003, for example, the Illinois legislature in the United States passed a law requiring that every “at-risk” child in the Chicago schools be enrolled in a “Scared Straight” Program, which would make the children visit prisons and meet with convicted criminals inside prison. This program looks very good on television, with apparently frightened 11-year-olds cowering in their chairs while criminals shout at them from behind bars. “Don’t be like me,” they say. “You don’t want to wind up in an awful place like this. It is terrible here, and only a very tough man like me can stand it.” While common sense may suggest that such sermons are enough to persuade any rational actor to obey the law, scientific research shows the opposite. In seven independent studies of this program in and outside the US, the net effect of the program was not to reduce the risk of crime but increase it. Anthony Petrosino and his colleagues (2003) report in a review of this research that children who were exposed to the Scared Straight program were more likely to be charged with crime in future years that children who were not exposed to the program, rather than less likely. So why, then, did the Illinois legislature enact the program?

The question of elected leaders ignoring scientific evidence is a large one, well beyond the scope of this article. It relates to many other questions of irrationality, including the matter of why doctors refuse to change medical practices after research finds them to be ineffective or life-
threatening (Millenson, 1997); the matter of why teachers refuse to adopt certain methods of teaching reading that produce far more students who can read than the teaching strategies teachers prefer to use (Mosteller and Boruch, 2002); or why many people continue to smoke cigarettes in the face of massive evidence that smoking shortens the life span.

The more basic question about crime prevention for any society is how it can generate the evidence it needs to decide what policies are rational, and what policies are irrational. Until such evidence is in hand, there is little point in persuading elected officials to endorse the principle that crime prevention policy should be evidence-based. What matters is the opportunity to engage in the practice of evidence-based crime prevention. Despite the growth of strong scientific evidence on crime prevention across the globe, most countries continue to suffer a complete vacuum of strong scientific tests of crime prevention methods. By the standards presented below, countries as advanced as Sweden and France, New Zealand and Japan, remain completely devoid of scientific tests of crime prevention policies. Only Australia, Switzerland, England, Germany and the US are known for certain (in English language reports) to have conducted scientifically controlled tests of crime prevention programs. Even in these nations, the number of such tests and the percentage of all programs tested remains minuscule.

If the first task for evidence-based crime prevention policy is to build a body of evidence, how can that be accomplished? Perhaps one way to think about this question is how not to do it. In May of 2003, the Minister of the Interior for Germany’s largest state, North Rhine-Westphalia, summoned the heads of all fifty police agencies serving the 18 million residents of the state. He announced to them that effective that day, all crime prevention programs in the state should be evaluated. In order to conduct such evaluations, all police agencies must immediately create evaluation units. The motto for this initiative was “Don’t just believe that it works, show that it works.”¹ No budget was provided for hiring scientifically trained program evaluators, no training was provided for existing police personnel, no standards, policies or even textbooks were provided to the German police chiefs sitting in the room. All they received was an introductory lecture from a bemused American criminologist, who could merely illustrate in English the kind of hard work they would need to do in German.

Another way, perhaps, not to develop a body of evidence is to rely solely on evaluations done in other countries. While such reviews as Sherman, et al (1997, 2002) may appear large-scale and comprehensive, they usually can say little or nothing about crime prevention programs outside of country in which they were written. The only way most nations can achieve political credibility for crime prevention evaluation research is to conduct such research in their own country. Even if studies already done abroad are only replicated within a nation such as Japan, they would answer the most important question on any reader’s mind: does it work here, in this culture and society? Since findings evaluating the same program in different parts of the same society can reach conflicting results (e.g. Sherman, 1992), it is all the more important to conduct crime prevention research in the country attempting to put such research to work. This may become more important from a scientific standpoint as the amount of unscientific and misleading evidence from

---

¹ I am indebted to John DiIulio for the idea of the German police initiative.
other nations appears on the world-wide web.

One solution to the challenge of knowing what research to believe from other countries, and hence what research to replicate in one’s own country, is the Campbell Collaboration. Thanks to the UK-based Cochrane Collaboration for systematically reviewing research on the effectiveness of medical treatments (see, e.g., Assendelft, et al, 2003), the International Campbell Collaboration was founded in February of 2000 to foster multi-national collaboration on reviewing research in crime prevention, education and social services. The Campbell Crime and Justice Group, chaired by Professor David Farrington of Cambridge University in England, has a website in English (http://www.aic.gov.au/campbellcj/) and, thanks to the efforts of Professor Hiroshi Tsutomi, in Japanese (http://fuji.u-shizuoka-ken.ac.jp/-campbellcj/index.html). This organization will grow in the numbers and scope of systematic reviews of crime prevention evaluations in years to come. But it can only grow as fast as the primary crime prevention research itself can grow. And for that, what is required is a commitment to doing such research in the first instance—preferably in the form of randomized controlled trials.

ANSWERING DIFFERENT QUESTIONS: The DRIVER Model

The recommendation to conduct randomized controlled trials (RCTs) should, in a purely scientific world, be almost completely without controversy. Since the late 1940s, RCTs have become the standard for scientific rigor in evaluating medical interventions (Piantadosi, 1997), and over one million RCTs have been published in medical journals (Millenson, 1997). They have also been widely recommended in crime prevention research, or at least endorsed as the most compelling evidence that can be generated on a research question of cause and effect (Sherman, et al, 1997). The merits of the RCT research design have been presented in treatises and textbooks in numerous statistical fields since the 1930s (Fisher, 1935; Cox, 1958; Campbell and Stanley 1963; Cook and Campbell, 1979; Federal Judicial Center, 1981; Boruch, 1997; Chalmers, 2003).

Despite endorsement by both academic authorities and the unrefuted mathematical logic of their ability to eliminate alternative, rival hypotheses, crime prevention officials in many countries remain reluctant to undertake RCT evaluations of crime prevention programs. Some of this reluctance is based on fear of the unknown. Some of it is based on unfounded attacks on the cost or complexity of such research. Little of it is based on the legitimate, if marginal, concerns of advanced statistical thinkers about how RCTs should be analyzed (Peto, et al, 1976; Weinstein and Levin, 1989; Gorman, 2002; Sherman and Strang, 2003) or what to do when the implementation of random assignment is less than perfect (Angrist, Imbens and Rubin, 1996).

Perhaps the fundamental barrier to implementing RCT research on crime prevention, however, is a mis-understanding of the questions they are designed to answer. The common mistake is to assume that RCTs are needed to answer all questions about crime. They are not! Rather, they are needed to answer the most important in a series of questions about how a democratic govern-

Evidence-Based Crime Prevention 85
ment can deal with emerging and changing crime problems over time. These questions have been taught at the University of Pennsylvania’s Jerry Lee Center for Criminology in the Fels Institute of Government (www.fels.upenn.edu) as the DRIVER model, which is an acronym that evokes a metaphor: government officials “driving” government like an automobile, for which evidence is providing the information on the dashboard: how fast the car is going, how much gasoline it has, how hot the engine is, and with computerized directional guidance, telling the driver which way to go next.

The acronym of “DRIVER” stands for

d-iagnostic data on distributions of problems over time and space
Example: hot spots of crime found in crime mapping (Sherman, et al, 1989)
R-espone to a problem with a policy based on reviews of research, preferably evaluations program options based on randomized trials.
Example: terminating “Scared Straight” programs in favor of “hot spot” police patrols which one Campbell review found to be effective (Braga, 2001).
I-mplementation of revised policy with reliable indicators of outputs
Example: weekly COMPSTAT meetings to examine crime patterns and police patrol allocations (Bratton, 1998)
V-alue added by the program to raw outcome measures based upon net impact of programs in relation to other known causes of outcomes
Example: estimating the effect of extra police patrols in hot spots based on comparisons to similar hot spots without police patrols.
E-valuation of the effectiveness of the crime-response policy as applied under local conditions, in light of local indicators and new social science evidence
Example: comparing the estimates of local spot checks of hot spots with and without police patrols to the estimates produced by previous RCTs (Braga, 2001).
R-evision of policy once again to take account of what has been learned over weeks, months or years of implementing and monitoring the delivery of the new policy.
Example: instituting new management systems to make sure police patrols are delivered in the time and place they are directed.

This model is an attempt to “manualize,” or make as explicit as an electrical appliance “user’s manual,” the steps needed to put research into practice-to undertake evidence-based government in response to such recurrent problems as crime and education. The DRIVER model draws attention to the Campbell Collaboration, and to the primary sources that Campbell employs, in order to reduce the use of misleading evidence and increase the awareness of cause and effect issues in all government programs, including crime prevention.

Teaching future government leaders that a systematic review is more useful than a single, selectively cited study may draw a clear line between training in law and training in government. All of the bias in presenting evidence that is accepted, indeed required, in the professions of advocacy
(like law) should banned in a curriculum for evidence-based government. The question of what can be concluded from the literature should not be a subjective one, although it can be. The question of what literature there is to analyze should become less subject to bias, and more comprehensively arrayed, with the aid of such growing sources as Campbell.

In addition to the accumulated literature of social science evaluations, the Fels curriculum builds on the growing use of the COMPSTAT ("computerized statistics") model in a wide range of government organizations. Pioneered in the New York City Police Department in 1994 as a means for reducing crime (Bratton, 1998), the model has since expanded into welfare administration in New York and all of city government in Baltimore (which calls the process "Citi-Stat"; see http://www.ci.baltimore.md.us/news/citistat/). The model is arguably a form of "evidence-based government," because it draws systematic attention to the trends in outcomes and other measurable aspects of public life that deserve a high level of attention. It is remarkable, in fact, that before the advent of post-mainframe information technologies, data such as crime trends were never examined on a "just-in-time" basis. Data were produced up to 11 months late, by which time little could be done with them.

The basic COMPSTAT model treats evidence with urgency, but not permanency. In a typical police COMPSTAT meeting at dawn, hundreds of officials gather in a large room with computerized crime maps, charts and graphs. The data under review cover the past weeks or months, rarely longer. The leading question is always what current efforts are under way to combat any increases or concentrations of crime. Unfortunately, COMPSTAT pays no attention to the research evidence on what crime prevention strategies may be most effective. In neither the US nor any other country reported to date is the first "R" of the DRIVER model part of the process. Instead the emphasis is placed on the "D" for Diagnosis: what, where, who, when and perhaps why, but not "works"—according to the research evidence from a rigorous controlled testm either here or in some other city. Such evidence is indeed permanent, and can be brought into the discussion on an urgent basis.

The confusion arises when a busy police executive may respond to such comments by saying "I don't need a two year randomized controlled trial to tell me where my robbery hot spots were last week." That statement is absolutely correct. It is also irrelevant. The "D" question of diagnosis can be answered with simple descriptive statistics. But the "R" question of what response to make to robbery, based on reviews of research done locally or elsewhere, may well require two-year RCTs to yield valid evidence. Doctor's do not like to guess when operating on patients with cancer. Why should police, or legislators, make guesses when "operating" on a city's rate of robbery?

In short, there are different kinds of evidence needed for evidence-based crime prevention, as follows:

**DIAGNOSIS** of crime problems: best done through comprehensive analyses of distributions of crime within an area by place, day of week, time of day, demographics of victims and offenders,
situational characteristics, weapons and modus operandi, etc. This kind of research is best known as **epidemiology**.

**RESPONSE** to crime problems based on research: best done through RCTs conducted in the local jurisdiction, for most directly applicable evidence; next best would be systematic reviews of RCTs; where RCTs are not available from any location, then reviews of quasi-experimental designs, featuring non-randomized control groups (see Campbell Crime and Justice Group websites referenced above).

**IMPLEMENTATION** of policies and programs requires operational data compiled by supervisors or service providers themselves, and have nothing to do with cause and effect or even epidemiology. The most relevant discipline here is **auditing**. The blunt question is whether people are doing what they are supposed to—like taking attendance in school. Do police provide patrol where and when it is assigned? Do nurses make home visits to all the new mothers they are supposed to train in how to raise a baby properly (Olds, et al, 1998).

**VALUE-ADDED** benefits of a program, as distinct from other possible causes of crime (from the weather, to economic trends, to competing or complementary governmental programs) can only be measured precisely through RCT designs. But they can be estimated in a less precise way as a local check on whether the results obtained are close to those reported in other RCTs, if any. This can be done by employing case-control methods of epidemiology (Piantadosi, 1997), as a quick backwards-looking check on whether there are differences in crime between units of comparable risk with and without the program—or with more of the program than with less of it (the latter variations can often occur locally through failures of implementation and management, which if measured well can provide a rich opportunity for Value-Added analysis). If no RCTs have ever been done, the value of conducting an RCT locally may depend on the amount of money spent on the program. The larger the cost of the program, the smaller the cost of the RCT as a ratio to program cost. RCTs are relatively fixed in cost and small as a percentage of programs. Thus for a program as massive as police patrol, or even airport security, the cost of RCT tests of the value-added benefit of different elements of a program is very tiny as a ratio of program costs. The same could be said for the testing of crime prevention programs for early childhood (e.g., Olds, et al, 1998). All of this can add up to the “profit or loss statement” for the chief executive: how much value (in crimes prevented) is returned on the investment of taxpayer dollars spent, according to our best estimates?

**EVALUATION** of a local program must meet a standard of assembling all relevant evidence, both locally and globally derived. The local evidence must start with the continuing accuracy of the **DIAGNOSIS**, as well as the ongoing success or failure of **IMPLEMENTATION** of the policy by local service providers. If nurses are being paid for home visits they simply are not performing,
then a local "evaluation" of the program could easily conclude the program is not being implemented, and is therefore wasting money. Alternatively, if the program is concentrating on low-risk mothers and ignoring high-risk mothers, then the evaluation may conclude that the program is serving the wrong client population and needs to be re-directed to "targets" of intervention that will yield more crime prevention per million Yen invested.

**REVISION** of programs at the local level, based on a local evaluation, must also depend on a wide range of data, including Diagnostic and Implementation evidence generated wholly in the local jurisdiction. The standard here requires much more than RCT evidence, which may or may not be useful—depending on the local issues. If the program has been properly implemented with the chosen targets, and the local "Value-Added" checks suggest that the return on investment is as expected, then there may be no need to revise a program. But if either targeting, implementation or value-added analysis indicates a concern, then some revision may be desirable based on all the evidence available.

**ARE RANDOMIZED CONTROLLED TRIALS NEEDED FOR ALL QUESTIONS?**

It should be clear from the analysis above that there is no need to use RCT evidence, or to conduct new RCTs, for many of the routine questions of government administration and crime prevention. All of the DRIVER questions require some kind of systematic and quantifiable evidence, the production of which is in some sense "scientific" (using replicable methods of observation and recording) and "research." Yet only two of the six DRIVER questions depend directly on RCT evidence. Thus the first point is that much, if not most, of the evidence needed in evidence-based crime prevention does not need to come from RCT designs.

The second point, however, is that there is a wide range of quality and potential corruption in non-RCT evidence. Crime data have been distorted and manipulated, early and often. The risk of under-counting criminal events for political purposes is very high in all democracies, as well as in dictatorships. The risk of distorting program data in order to make a boss look good is just as great. Thus the discipline of auditing, with independent external examination of program claims, is highly relevant to much of the non-RCT evidence in crime prevention administration.

The third and most important point is this: RCT evidence is the single most important kind of evidence to obtain in government administration, including crime prevention. Regardless of how small a percentage it is of all the evidence needed, it is absolutely necessary to prevent all other evidence from becoming irrelevant. Imagine, for example, a perfect system of diagnosis, implementation, value-added analysis, evaluation and revision. Now imagine that all of this work is undertaken without the benefit of a single RCT testing the major premise of the program—that it works! If the program were something like "Scared Straight," which causes rather than preventing crime, it is entirely possible that the program could look like it is doing good while it is really
doing harm. Only the addition of the RCT design to the DRIVER model at the “D” and “E” phases can prevent such tragic mistakes.

The answer to the question of whether RCT evidence is needed to answer all questions is no. The answer to whether RCTs are needed to answer the single most important question is yes. This answer has major implications for crime prevention policy in all nations.

ALL NATIONS NEED THEIR OWN RCTS FOR CRIME PREVENTION

In a recent test of restorative justice in Australia, Sherman, Strang and Woods (2000) found that face-to-face meetings of victims and offenders in violent crimes caused a substantial reduction in repeat offending compared to a control group. In contrast, the method did not work for property crime. When the authors replicated their work in a second English-speaking country, however, they found exactly the opposite: that the method of restorative justice meetings reduced repeat offending after property crime, but did not work for violent crime (Sherman, et al, in progress). Close inspection of the data suggests that the difference in results in unlikely to be caused by the flaws in the respective RCTs. Rather, the difference appears to be that the program had different effects on crime in different social settings. While the exact nature of those differences can only be determined by conducting many more such RCTs, the conclusion for each country can be reached immediately: rely more on RCTs closer to home, or more similar in social context.

Now suppose Japan had adopted restorative justice for juvenile violence based on the Australian evidence alone. It is possible that in Japan, juveniles respond to restorative justice just as they do in Australia. Thus the program may have caused a reduction in crime. But it is also possible that juveniles in Japan are more like juvenile offenders in the second nation to conduct an RCT on restorative justice for violence. In that nation, preliminary evidence suggests that it increases repeat offending for boys while reducing it for girls. So what is the effect of the program in Japan? It seems unlikely that evidence from other countries will answer the question. If Japanese leaders want to know whether to adopt restorative justice conferences for violent crime by Japanese children, there is no better scientific course available than to conduct an RCT of this program in Japan, with Japanese offenders, victims and police.

The implication for RCT policy could not be clearer. As valuable as the Campbell Collaboration is, no nation can rely on RCTs done in other nations to know about what works in what country. Major nations like Japan, blessed with centralized governments and large budgets, are in a prime opportunity to conduct their own RCTs—more so than, say, New Zealand, with a population a tiny fraction of Japan’s. Said a different way, the fractional cost per crime prevention program budget of conducting RCTs on each program is far lower in Japan than in most other countries. This means that the benefits of administering crime prevention with RCT evidence can far outweigh the costs of undertaking them. It also means that the cost of wasting money on ineffective crime prevention programs is far higher in Japan in elsewhere, because the relative cost
of detecting program failure with RCT evidence is so much lower in Japan.

These are the utilitarian reasons for Japan adopting a policy of RCT testing of all major crime prevention programs, or at least testing new ideas before major expenditures are invested in them---no matter how severe the crime “crisis” caused by the most recent tragic event. Utilitarian reasons, however, may be less inspiring than reasons of higher value. Evidence-based policy is a higher value than emotion-based policy, or superstition-based policy, or prejudice-based policy. If democracies can agree on goals, based on human rights (such as freedom from crime and freedom from injustice), then the world in which evidence-based policy guides crime prevention is a better world than one in which emotional responses guide policy (Sherman, 2003).

The most important reason, then, for Japan and other nations to conduct RCTs on crime prevention policies is to inspire even more nations to do the same. Nothing spurs adoption of new ideas like other actors doing the same. Some actors are early adopters, while most are later adopters. The leadership from early adopters clearly influences the later adopters, in things both bad and good (Gladwell, 2000). If we agree that evidence-based crime prevention is a good thing, or at least better than its alternatives, then leadership can become a very inspiring reason to adopt an RCT strategy. Leading by example in evidence-based crime prevention is the challenge and opportunity facing Japan, and all nations.

Notes
1 Speech by Fritz Behrens, Minister of the Interior, North Rhine-Westphalia, Germany, to police executives at the Polizei Fortbildungsanstalt, Neuss, May 19, 2003.

References


E-mail: lws@sas.upenn.edu
エビデンスに基づいた犯罪予防
——国際的視点（米国から日本へ）——

ローレンス W.シャーマン
ペンシルバニア大学

民主政治において、新たな犯罪の原因を作らずに犯罪を防止するにはどうしたらよいのであろうか。耳に心地よい反面、効果のないプログラムに予算を浪費しないためにはどうしたらよいのであろうか。そして、情動ではなく、理性に基づいた決断をするにはどうしたらよいのであろうか。こうした問題は、文化や歴史にかかわらず、どのような民主体制をも悩ませる問題である。これらについて、エビデンスに基づいた政府においては共通の解決策があるかもしれない。この論文は、エビデンスに基づいた政府という考え方及びそれを実践するためのDIRVERモデルを見つめ直すと同時に、犯罪防止に関する様々な疑問に対するエビデンスの基準についても考察する。そうした考察に基づいて、犯罪防止のエビデンスにおいて、無作為統制実験（RCT）が必要とされる部分はそれほど多くないかもしれないが、新たな犯罪を生み出すことなく、犯罪を防止するためには、それぞれの国においてRCTが不可欠である。犯罪防止政策をより理性的な枠組みで考えるためにも、先進国の指導者は、RCTを推進していくことが必要である。

キーワード：エビデンス、実験、犯罪予防

E-mail: lws@sas.upenn.edu