Inter-American
Development Bank

Institutions for
Development
(IFD)

TECHNICAL NOTE

No. IDB-TN-436

# Developing and Evaluating Citizen Security Programs in Latin America

## A Protocol for Evidence-Based Crime Prevention

Lawrence W. Sherman

Cambridge University and
University of Maryland

July 2012

# Developing and Evaluating Citizen Security Programs in Latin America

## A Protocol for Evidence-Based Crime Prevention

Lawrence W. Sherman

Cambridge University and
University of Maryland


Inter-American Development Bank

2012

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Sherman, Lawrence W.
Developing and evaluating citizen security programs in Latin America : A protocol for evidence-based crime prevention / Lawrence W. Sherman.
p. cm. — (IDB Technical Note ; 436)
Includes bibliographic references.
1. Crime prevention—Latin America. 2. Public safety—Latin America. I. Inter-American Development Bank. Institutions for Development Sector. IV. Title. V. Series.
IDB-TN-436

http://www.iadb.org



Contact: Lawrence.Sherman@crim.cam.ac.uk
+1-267-269-1757

## Abstract

This protocol is designed mainly for people working to reduce crime and improve justice in Latin America, but it discusses principles that can be used anywhere in the world. Those principles can be summarized as *evidence-based crime prevention,* a process by which good evidence on the facts of crime and its prevention is at the heart of theories and programs for promoting citizen security. "Evidence" in this sense is broadly defined as systematic factual observations of all kinds, not just as the forensic details of a criminal case. Evidence is the data developed by scientific methods to observe and predict any kind of truth, including facts about health, education, crime, and justice. A few key principles of evidence-based crime prevention are the following: crime must be *measured* reliably and precisely by well-audited systems; crime should be *classified* in whatever way supports crime prevention; and crime should be *analyzed* in multiple units or categories, including offenders, criminal networks, victims, micro-places (hot spots), communities, times, days of the week, and other categories.



**JEL Classification:** Z18

**Keywords:** impact evaluation, citizen security, cost-benefit, evidence-based crime prevention, efficiency and effectiveness

# Table of Contents

# A Framework for Planning Innovations for Citizen Security

## Introduction

Where do good ideas come from? Making Latin America safer certainly requires good ideas. The question of how to find them must be the starting point for any effort to increase citizen security (Johnson, 2010). That question quickly leads to its companion: how to distinguish good ideas from not-so-good (or bad) ideas (Sherman, 2011a). It is only by answering those two questions that we can answer a third, more vital question: *how do we put good ideas to work to build safer societies?*

Some say good ideas come from inspiration or even intuition. Others say good ideas come from systematic analysis. According to Nobel Prize winner Daniel Kahneman (2011), both views are right. People think in both ways, often at the same time. One way of thinking is fast and the other slow and they use different parts of the brain—"System I and System II," as psychologists describe it. System I is automatic and easy, with no sense of controlling the thought process. System II is difficult and methodical, requiring concentration and, often, computation. Falling in love is done with System I; building a cathedral is done with System II. But love can also inspire good ideas for building a cathedral.

For the past 500 years in the West, and even earlier in Asia, thinkers have tried to separate the two ways of thinking. From the works of Francis Bacon and René Descartes to Malcolm Gladwell's *Blink* (2005), the merits of methodical thinking have been praised over the dangers of automatic thinking. As Gladwell suggests, sometimes a snap judgment produces better results than a careful decision. But he also documents the many disasters that have resulted from relying solely on System I, with too little attention to thinking with System II. Good results for complex problems depend on both good judgments based on experience and careful analysis of facts and theories of causation.

## Purpose

This protocol provides a framework for improving citizen security with System II thinking. Put another way, it is designed to supplement the good judgment of experienced leaders with

systematic analysis to better support that judgment. Its method is well-tested, and hardly original in concept. What is original is the effort to link citizen security practices to one of the most successful inventions in human history: science.

Scientific methods do not require laboratory coats, microscopes, or test tubes. Science is simply a systematic way of organizing evidence in relation to any claim about how the world works and reaching conclusions based on that evidence. It is by no means an uncontroversial idea. Science has been both widely adopted and widely attacked, both praised and punished, depending on the whims of those in power. Democracies and dictatorships alike have refused to accept overwhelming evidence that ran against conventional wisdom or "blink" snap judgments. The prominence of intuitive thinking in civic voting has deterred even the most enlightened leaders from relying on good evidence. Yet when it comes to crime prevention, any society can conclude in a blink that traditional methods have not been good enough; new and different methods must be tried as innovations.

So what does this protocol add to the extensive knowledge (and strong emotions) many citizens and leaders already have about innovations in crime prevention and justice? The answer can be summarized in one word: testing. An idea cannot prove itself by logic or instinct. Good ideas can be separated from bad ideas only with tests in the real world.

Most innovations in crime prevention are tried but not tested (Sherman, 2011a). That is, people who have faith in the innovation want to try it by doing, but refuse to test it by direct comparisons. Comparing the new idea to an old idea, they say, is unethical, because they are already sure that the new idea works better. This argument is often made for new "miracle cures" in medicine (Millenson, 1997), only to discover that they result in such tragedies as children born without limbs (from thalidomide) or enormous wastes of money (such as from tonsillectomies). In the long march of medicine, the good judgment of experienced doctors has been increasingly supplemented by legal requirements that all drugs be *tested* for safety and effectiveness before they can be put on the market.

A true test must be designed in a way that the innovation can fail the test. Most often, however, innovations are "doomed to success," as English police leaders say with wry humor. Trying without testing is anti-scientific, since true tests must clearly state *how* any hypothesis can be falsified (Popper, 1959). A nonfalsifiable hypothesis is what faith, hope, or other good

things depend on. But it is not the way humans have generally solved complex problems like crime.

This protocol sets out steps anyone can take to develop and test innovations for promoting citizen security, using the best evidence available. It includes the use of theory in explaining *how* and *why* innovations work or don't work. It includes the roles of implementation and measures of delivery of tasks on which better outcomes depend. Most of all, it shows how to use the principle of *fair comparisons* in reaching conclusions from evidence.

The protocol uses the acronym DRIVER to mark the six steps of evidence-based innovation:

D for diagnosing the nature and direct causes of the crime problem

R for revising current practices for dealing with it

I for implementing the revised practices as they were designed

V for the value-added cost-benefit theory of the innovation

E for evaluation of the implementation and theory of the innovation

R for reviewing more ways in which practices could be revised and tested

The DRIVER steps are described in detail, first in the next section and later in a protocol checklist for each step. To appreciate their logic, it is useful to build on the foundation of the 10 principles listed in the Executive Summary and described fully below.

**10 Principles of Evidence-Based Crime Prevention**

1. **Crime must be *measured* reliably and precisely by well-audited systems.** Good measurement is as important for dealing with crime as it is for weather forecasts, food production, economic growth, and many other issues in complex societies. Over the past two centuries, two major systems for counting crime have been developed. One is counting crimes reported to the police. The other, more recent method is using crime victimization surveys.

   Counts of crimes reported to police are always incomplete because of citizen decisions not to tell police about some crimes. Police may also under-count crimes owing to political pressure to make crime rates appear lower than they really are. Victimization surveys provide

a valuable check on both those threats to valid crime measurement. But they cannot take the place of police gathering detailed facts about each crime, including the exact location, date, and time of each criminal event.

The safest nations on earth, from Japan to Sweden, place great emphasis on police recording the details of all crimes reported to them in digital formats that make these reports more useful for crime analysis. The United Kingdom has placed even greater emphasis on auditing police crime reporting, with dedicated positions for crime "registrars" who regularly sample crime recording in police stations. Yet the United Kingdom has yet to provide accurate data on the location of each crime; a popular website (www.police.uk) that claims to do so contains massive errors in the crime maps (Palmer and Caddick, 2010).

The most advanced technology for pinpointing crime locations comes from global positioning satellite (GPS) systems, which are already in use by police in some developing nations (such as India) but not in most of the G20 countries. Adopting both GPS and the portable cameras with microphones that police can now wear to record interviews with crime victims and witnesses can free police from ever making crime reports again. Instead, the digital record of their video and audio input for each shift of duty can be downloaded and sent to trained crime registrars, who can fast-forward through the video records to find and code reports of crime taken by officers with the help of a checklist that prompts officers during interviews.

2. **Crime should be *classified* in whatever way supports crime prevention.** The most important feature of crime measurement is the use of incident-based reporting systems (IBRSs) rather than merely categorical totals by police area. Pioneered in the United States by the FBI in the 1980s, IBRSs provide multiple options for classifying and analyzing crimes. This flexibility allows police and policy-makers to adapt to changing crime patterns by reorganizing the way they are measured to fit operational concerns. That stands in sharp contrast to the traditional adherence to the *Penal Code* as the only system of classification. It also provides the capacity for timely and accurate intelligence about crime that has been the hallmark of evidence-based police initiatives such as COMPSTAT in New York City and elsewhere (Bratton, 1998).

Why is it unwise to rely solely on a penal code to classify criminal events? The answer lies in the broad range of facts covered by any legal category. In the case of murder, for example, many different weapons are used to kill people. Most uses of these weapons, however, do not result in death. Hence the *behavior* of shooting at someone may be classified under different crime categories depending on whether the bullet misses the person shot at, hits the person but does not kill, or causes a fatal injury. Penal codes place shootings that do not result in death in the same legal category as fistfights and stabbings. There is no way to track trends in total crimes with guns by relying solely on most penal codes. Yet guns may be (and often are) the crucial issue in reducing harm from violent crime, as research in Bogota and Cali has demonstrated (Villaveces et al., 2000).

Problem-oriented policing (Goldstein, 1990; Braga and Bond, 2008) and other environmental strategies for crime prevention may look for even more detailed information about criminal events. These details may include lighting conditions, proximity to taverns and the hours that taverns are open, location of the nearest highway, and weather conditions. None of these characteristics can be used to classify crime in ways to suggest better prevention practices unless a computerized record is created for each criminal incident. This is not bureaucracy; an IBRS is the basic source of intelligence for strategic action.

3. **Crime should be *analyzed* in multiple units, or categories, including offenders, criminal networks, victims, micro-places (hot spots), communities, times, days of the week, and other categories.** Once incident-based data are available, they can be sorted and sliced according to many different ways of targeting crime prevention practices. One common way to do this is by individual offender. But few offenders commit crimes on their own. Most crimes are committed with co-offenders or by offenders who are part of social networks of criminal activity. Sometimes those networks are called “gangs”; more often, they are unstructured waysof intermittent contact, much like reading Facebook pages. These networks may also include crime victims, especially young males who are at the greatest risk of violent injury or death. Other victims may be innocent of any social connection to their offenders, but may be repeatedly targeted in ways that can be dealt with by crime prevention.

   There is growing evidence that the traditional focus on *people* may be less useful in crime prevention than a focus on *places*. Concentrations of crime across micro-places, or “hot

spots" of crime (Sherman, Gartin, and Buerger, 1989), appear to be more intense than across offenders (Weisburd, Sherman, and Maher, 1993). Analyzing crime by smaller geographic units can reveal that most places in even high-crime communities are crime-free. This means that resources, such as those used for patrolling, may be wasted on streets that have had no murders in 10 years. A recent micro-local analysis of murder in London over 10 years found that 42 percent of the thousands of murders were committed in just 6 percent of the many more thousands of small area units (Jackson, 2011). Similarly, Weisburd et al. (1993) reported that just 5 percent of the street segments in Seattle were the locations of over half the crime for over 15 years.

These high concentrations of criminal events in certain places can interact with concentrations among individuals. Wikstrom et al. (2010) found that even high-frequency delinquents were many times more likely to commit crimes in a crime hot spot than in a low-crime location. Taking these facts into the realm of time, it can be readily predicted what days of the week and times of day have the highest risk for crimes committed by the highest-risk offenders against the highest-risk victims.

For certain kinds of crimes, however, offenders may remain the most useful subject for analysis. Domestic violence is a prime example of the need for offender-specific databases, although victim-based data can also be useful. The value of both kinds of data is in its ability to build prediction models, helping police and officials to set priorities for the offenders or victims most likely to be involved with homicide or serious injury.

4. ***Priorities* for crime prevention should be set within types of units based on the degree of *harm* caused by different kinds of crime.** Limited budgets will always require choices to be made about how to focus resources. These choices form priorities that may often be distorted by lumping all crime types together. Many governments, including those of the United States and the United Kingdom, report the nonsensical figure of "total crimes" reported each year. By this metric, a homicide is the same as a purse-snatching; a rape is the same as shoplifting. No penal code suggests such utilities. Sentencing guidelines in many countries now indicate very precise recommendations for how many days of prison time—including zero—should be the starting point for any sentencing decision. These counts of days or months in prison provide a reasonable basis for calibrating the harm caused by each crime (Sherman, 2010).

The harm caused by crime is a rational System II concept that blends nicely with emotional System I political viewpoints. The recommended length of prison sentence creates a currency of harm assessment that can be used to set transparent priorities for crime prevention investments. It can also provide a clear basis for reallocating the "stock portfolio" of such investments, such as between prison and probation, corrections and policing, or early childhood risk prevention versus adult rehabilitation.

This approach can be made transparent with a Crime Harm Index (CHI) that multiplies the level of harm each crime causes by the number of crimes of that type. A CHI can replace a confusing array of crime statistics, with some crime types rising while other types fall. Instead, the CHI offers a single bottom-line calculation of whether crime has caused more harm this year than last, in this community than in that one, or in this group of ex-prisoners than in that one. Impact evaluations of crime prevention strategies can be made far more informative by using a CHI rather than such crude measures as percentage of offenders reconvicted or rearrested (recidivism).

A CHI-based outcome provides the only real equivalent to profit-and-loss analysis in a business environment. While businesses always report earnings in a common currency, criminal justice does not; each crime type is, in effect, a different currency. No one would add together pesos and euros and dollars to measure how many pieces of "money" they have; they would try to peg all their cash to a single currency that can be equated to any other single currency. That is exactly what a CHI can bring to crime analysis and prevention.

5. **A *power few*, or small proportion, of all units of criminal behavior causes most of the harm for most crime types.** No matter how the harm of crime is calibrated, it is highly concentrated in a small tail of any distribution (Sherman, 2007). Whether offenders or victims, places or times of day, the risk of future crime varies widely. It does not fall on a bell curve (normal distribution) across all units at risk. Virtually every analysis of crime by any unit shows a *power curve,* the hockey-stick shape first identified by the early economist Vilfredo Pareto to describe wealth and earnings.

This distribution is both obvious and potentially inflammatory. It is obvious, for example, that high-crime areas experience far more crime harm than other areas do. Yet it is inflammatory to propose that low-crime areas should receive less attention than high-crime areas. Taxpayers in low-crime areas may pay more taxes, and thus demand more protection. It is hard to explain that they are most at risk, wherever they may go in their society, when the high-crime areas are left with insufficient crime prevention investment.

A focus on the power few can have benefits far beyond the immediate areas where the few may live or commit crimes. An entire society can benefit from reducing the violence in its epicenters, even without attention to the low-crime periphery. Explaining this to the broader population requires courageous and skilled leadership. It is at once a matter of System II thinking that can be made part of a System I intuitive understanding: "Of course we need more police where the shootings are most common!" The obvious character of such statements may well counteract the obvious implication of robbing Peter to pay Paul, or reallocating taxes according to need. Perhaps the best points for such explanations are efficiency and effectiveness.

6. **The *efficiency* of crime prevention can be greatest when resources are concentrated on the power few units, identified using predictions derived from their past behavior.** This principle is derived from pure logic. Like many logical arguments, it depends on a key assumption about empirical facts. The assumption is this: When the same practice reduces harm from crime by the same proportion across all units for the same cost, the units with the most crime can yield the greatest absolute return on investment in reduction of crime harm. For example, if it costs $50 per hour to have a police car patrol one area, and crime is 50 percent lower when police patrol than when they do not, there will be more crimes (and harm) prevented per dollar in a high-crime area than in a low-crime area. That, at least, is the assumption.

   The evidence for this assumption is not strong, but it is consistent. Ratcliffe and colleagues (2011) showed that the same amount of foot patrolling per square meter in Philadelphia reduced crime most in the highest-crime foot beats, with less crime reduction as crime levels declined. Yet this isolated finding is limited by small sample sizes and lacks replication. The breadth of Principle 6 also encompasses far more than police patrol. But at

least a similar finding was reported by Olds and colleagues (1997) for a program of highly trained nurses visiting the babies of new mothers, with the most cost-effective reductions in child abuse found among those at highest risk. A program of restorative justice for juvenile delinquents in Australia found the greatest reduction in recidivism among the most frequent offenders (Woods, 2010).

Further support for this principle (and its key assumption) can come from a systematic review of all possible evidence. Until such a review is completed, both the theory and facts support the efficiency of concentrating on the power few.

7. **The *effectiveness* of crime prevention is best determined by field tests that compare crime rates among people, areas, or other units under different programs.** The distinction between efficiency and effectiveness is crucial. Efficiency is possible only when practices are effective—that is, when they produce the results (or outcomes) they were intended to achieve, such as preventing crime. An ineffective program can never be efficient. An effective program, however, can still be inefficiently allocated. It may not even "work" for some kinds of people, or cause more crime for some and less crime for others (Sherman, 1992). The only way to determine the effectiveness of crime prevention is to put it to the test, in real-world conditions.

   What constitutes a test? The best definition of a test is a comparison of crime outcomes across similar units, with and without the use of the practice being tested. The units, once again, can be offenders, victims, places, or days of the week. The practice tested can be compared with another practice or with doing nothing. The use of a control group in an experiment is entirely for the comparison of outcomes. To say that an impact evaluation shows that a practice works requires a comparison by definition: it works compared with what?

   The choice of comparison is a vital part of developing crime prevention ideas. For example, when electronic monitoring of offenders on probation in the United States was compared with not monitoring probationers, the findings may show little benefit of monitoring (Renzema and Mayo-Wilson, 2005). But when electronic monitoring was compared with prison sentencing in Buenos Aires, the results showed that monitoring produced less crime at less cost (Di Tella and Schargrodsky, 2010).

The clearest tests of crime prevention practices compare them with doing nothing at all. The clarity comes from knowing exactly what it costs to deliver the practice being tested, with the cost of the comparator clearly being zero. If nothing has been done with those units in the past, the ethics of doing nothing may be more acceptable. If limited resources are available for a new idea, there will have to be some sort of selection process in any event. The principles of "need" versus "evidence" will then collide, but only in the short run. Giving a program to all of those in greatest need may guarantee that the program will die for lack of good evidence of its effectiveness. Giving it to half of a larger population with high need can produce good evidence that the program is cost-effective. In the long run, everyone benefits from knowing what works and what doesn't. At the very least, such controlled testing can discover whether a program that looks good in theory is actually harmful in practice (McCord, 1978, and McCord, 2003).

Comparisons with other practices can also produce clear tests. The clarity can come from unbiased selection of each unit for one practice or another. When the Minneapolis Police Department randomly assigned domestic violence suspects to be arrested or counseled, assigning the suspects at random to the two different practices helped to reduce any bias in the comparison of recidivism between two groups of 100+ suspects (Sherman and Berk, 1984).

Such a bias could come in the form of, for example, choosing to arrest suspects who are less polite to police, and perhaps more likely to repeat their alleged violence. Or it could be arresting more people who had previous criminal records than people who had no prior arrests. Random assignment in that experiment was not perfect, but modern technology now makes it possible to reduce bias far more effectively through Internet-based platforms for random assignment (Ariel and Sherman, 2012).

Tests by comparison, using random assignment, are usually called experimental; tests with comparisons not based on random assignment are usually described as quasi-experimental. These test designs have been ranked from 1 to 5, with 5 indicating a random assignment experiment and 1 meaning an observed correlation. A level 2 design is simply a before-and-after comparison of the same group of units, with the crime outcomes from before

the introduction of a practice compared with the same outcomes after the practice is introduced.

What is wrong with a before-and-after comparison? The problem is that it is uncontrolled. There could be many reasons for the outcomes to change, or not, other than the practices being tested. The lack of a comparison, or control, group means that there is no way to estimate what might have happened without the introduction of the new practice, since crime trends often fluctuate naturally or by chance. A before-and-after comparison can appear to show that a program did not work when in fact it did—or that it did work when it really did not. Far too many multi-million-dollar decisions are based on these tests, which would never be allowed in health or agriculture. Crime is important enough to merit a higher standard of testing than before-and-after comparisons (Campbell and Stanley, 1966).

Three examples of citizen security evaluations in Latin American countries illustrate, in different ways, the value of using comparisons. One is the Villaveces et al. (2000) impact evaluation of a police program to ban gun-carrying, in cars or on the person, on holiday weekends in Bogota and Cali. Because gun homicides had a history of sudden spikes on holiday weekends, medical staff who dealt with shootings had an interest in seeing whether the police action could reduce injury. The pro-action consisted of mounting roadblocks where cars and people could be checked for concealed weapons, which could theoretically work as follows: Gun checks lead to less gun-carrying, which leads to less gun density in public places, which in turn leads to less impulsive shooting because fewer guns are available. By comparing violence on holiday weekends before and after the program began, and by comparing other holiday weekends when the program was not in operation, the authors were able to estimate the modest but significant impact of the program in preventing homicides.

A second example of a comparison comes very close to the ideal of a randomized controlled trial. Di Tella and Schargrodsky (2010) compared cases of serious criminals being sentenced in Buenos Aires by judges who had differing preferences for how to punish the crimes. Some judges preferred to send offenders to prisons while others preferred to sentence them to house arrest with electronic monitoring. Because the cases were assigned to the different judges by a random number formula, the other possible causes of any difference in repeat offending would be likely to affect both groups equally—and hence control for those

differences. The authors observed that offenders sentenced to house arrest actually had lower rates of repeat offending than those sentenced to prison (a finding consistent with other studies since then showing that prison sentences increase offending rates more than not sentencing similar kinds of offenders to prison). This evidence is very strong and useful for developing sentencing policy in Latin American countries.

A third example is less rigorous, but also makes good use of controlled comparisons. Soares and Viveiros (2010) evaluated the adoption of a COMPSTAT-like integration strategy for different police forces in Minas Gerais, Brazil. The police management program was modeled on the COMPSTAT system developed by the New York Police Department in the 1990s (Bratton, 1998). The authors compared crime rates in the areas in which Minas Gerais police had adopted the system to areas in which they had not. By using a before-and-after comparison between areas adopting the system and areas not adopting it over the same periods, the authors were able to conclude that the program reduced property crime but not homicide.

Yet the third example is also a useful lesson in how not to evaluate a program. Any evaluation that is forced to make comparisons after the fact is limited in the kinds of differences that can be compared. It is hard to be sure that apples are being compared with apples rather than oranges; that is, that comparisons of areas with and without the programs are made using units that have similar crime problems, demography, and so on. The authors did the best they could with what they had. But when program development and evaluation work hand in hand, there can be much better matching of experimental and comparison units. There can also be attention to such details as making sure the comparison areas do not lie adjacent to the experimental areas, hence being subject to potential crime displacement or spillover. But even then, there are ways to make random assignment make even modest evaluations more rigorously controlled.

Even when a small number of units are available, there are ways to increase the validity of estimated causal effects. Using these principles allows planned studies to make much more reliable comparisons, even if they cannot rise to the level of a randomized trial. There are always reasons why such highly controlled tests may not be possible. When areas or other large units are the subject of a test, for example, there may not be a large enough number of them to enable overall random assignment to remove biases. In those cases, 20 or 30 units

can be paired as closely as possible. Then within each pair, the practice being tested can be randomly assigned to one member of the pair but not the other (see, for example, Braga and Bond, 2008). The average difference in outcomes within each pair can then be computed as the effect of the practice on crime, relative to the comparison practice (or no practice).

Random assignment experiments are often denied on the grounds of cost. Recent evidence in the United Kingdom and the United States, however, shows that good random assignment tests can be done more cheaply than other kinds of evaluations (Sherman, Strang, and Ariel, 2011) and produce more rigorous evidence. This also means they are better able to explain the results, especially in reference to the theory of how and why the practice works—or doesn't.

Finally, the basis of any test is a clear conception of the goal to be accomplished by a crime prevention practice or program. Penicillin was never tested to cure polio, diabetes, or cancer; the focus for it was on infections. Yet many crime prevention initiatives are justified on the grounds of preventing many kinds of crimes, with no clear theory of how they work.

8. **The *theory* of any crime prevention practice should specify a logical sequence by which the practice produces measureable *outputs* of action that reduce *outcomes* of crime or injustice, explaining *how* and *why* the practice should prevent crime.** People are often skeptical when told that something works to prevent crime (or doesn't), even with evidence from good experiments. Before they accept the conclusion, they want to know why the thing works. The truth is that experiments, like football teams, are not always able to offer plausible reasons for their results; they simply report results honestly for others to interpret. Big, satisfying theories like "leadership" or "team commitment" are hard to test in the context of either sports or impact evaluations.

   What every experiment can do, however, is to spell out who needs to do what—just like a football team does. If the goalie fails to block the ball, the team will lose. If the police fail to perform patrols, the experiment will not show a crime reduction. Every crime prevention practice needs to distinguish between *outputs* of actions certain people are supposed to perform and the *outcomes* intended to result from those outputs. More importantly, both outputs and outcomes should be measureable.

Outputs are easier to measure than outcomes. Outputs can be measured largely by counting the units to be delivered. In an experiment that increased patrols at some crime hot spots (and not others), graduate students with stop watches counted how many minutes police patrolled at each of 100 hot spots (Sherman and Weisburd, 1995). This observation allowed the experiment to report the ratio of police patrol minutes in the experimental group to minutes in the control group. This ratio was the prime indicator of the level of police patrol output, which was hypothesized to affect the crime rate (it did). With twice as much patrolling in the 50 experimental group hot spots as in the 50 control group locations, the experiment revealed a clear difference in dosage of police patrol. The difference in crime could then be shown to follow the difference in patrolling.

Today, such labor-intensive measurement is not necessary. Police officers carrying portable radios usually have GPS devices embedded in the radio. This passive measurement of their whereabouts at all times allows a single data analyst to reconstruct the number of patrol minutes at every location in the experiment at very low cost. What is essential for any experiment, however, is to measure the difference in output between the target and comparison groups.

Sometimes outputs are multiple and sequential, with each step requiring separate measurement. In a sports-based after-school program, for example, the first output to measure is the opening hours of the program—not the *scheduled* hours, but the actual hours of operation independently measured by someone not affiliated with the delivery of the program, or by a passive technology such as closed-circuit television. The second output could be enrollment in the program: how many children have signed up to attend the program five days a week? The third output could be actual attendance at the program: how many children, on average, actually attended the sports program each day? How long did they stay?

The outputs of the sports-program crime prevention initiative could be consistent with the criminological theory of "routine activities" (Cohen and Felson, 1979). The more time high-risk children spend under adult supervision, the lower the probability that they will commit crimes or have crimes committed against them. This theory depends on good measures of how children spend their time. This is not bureaucracy; this is vital evidence. If any of the links in the following chain are broken, then there is no reason to expect that the program will prevent crime:

sports center open → children enrolled→children attend→fewer opportunities to commit crime →less crime by or against children.

Note that this theory is all about the children enrolled in the sports center. It is not about the homicide rate in the city or country where the sports center is located. People who propose such activities often suggest a wide disconnection between the delivery of the program and a likely outcome of the program. Spelling out the theory, with all its outputs, is a good way to keep on track with a theory of how and why the program should change crime rates. Measuring the outputs can reveal whether the *theory* itself was wrong or whether the delivery of the outputs simply did not happen (or did not happen to the threshold of a minimal level). It also keeps the outcomes closely tied to the theory of how the outputs change outcomes, rather than having the outcomes measured far away from the outputs.

In the sports center theory, the unmeasured link is the number of criminal opportunities. This is a prime example of a "black box" of causation. If all the outputs are delivered, we still may not know whether the children enrolled in the program had fewer opportunities for crime. If the program then fails to prevent crime, it could be for that reason; it could instead be because the program simply did not work as theorized. Whether or not such key links are measured, any connection between delivered outputs and differences in crime outcomes remains an important contribution to knowledge. Evidence-based crime prevention does not let the perfect be the enemy of the good. If most outputs are measured, their delivery is enough to test the basic form of the theory.

9. **The *ethics* of a crime prevention practice depend not on its success in reducing crime, but on whether the practice a) respects human rights and b) maintains a proportionate balance between the harms of coercion and the harms of crime.** This is a principle of all efforts to prevent crime, public and private. It is essential to keep this principle prominent in the transparently Utilitarian framework (Beccaria, 1764; Bentham, 1781) offered by this protocol. Even if practices are effective by some measures, they cannot be acceptable if they violate human rights or fundamental ethics. Some police think the torture of criminal suspects is an effective way to solve crime, but that hypothesis is ethically untestable.

Introducing alternatives to torture, such as DNA collection at crime scenes, could be an ethically acceptable way to try to reduce torture. But the distinction between what works and what's ethical must remain a bright line.

What is less clear is the tightrope governments walk over what is often called the balance of harms. Many criminal justice practices intrude on liberty in ways that are known to be harmful, such as arrest and imprisonment. Police stop-and-search methods are a prime example: innocent people are stopped in high-crime areas to prevent deaths from gunshot wounds (Villaveces et al., 2000). Such stops are manifestly intrusive. Whether they are ethical, however, may depend entirely on the cost-benefit ratio they produce (Sherman and Rogan, 1995). If lives can be saved by conducting such police operations, without injury or lasting harm to anyone, many courts and voters would hold the operations to be ethical. Others might disagree. The line between right and wrong cannot be bright in such circumstances. But it can be made brighter with the evidence from good impact evaluations.

10. **The best evidence for developing and improving local crime prevention practices draws on both international impact evaluations and on local crime analyses of all kinds, including assessments of past efforts and predictions of future crime patterns.** Where do good ideas come from? They come from anywhere and everywhere, globally and locally, from System I blinking and System II thinking. The global scientific literature on crime prevention is vast and grows larger every day. Flashes of insight and good ideas can also emerge every day, but in less accessible or transparent ways. Most important may be systematic evidence on local crime problems, which indicates the main priorities and targets of opportunity for reducing the harm of crime.

    Many innovations in crime prevention are launched without anyone ever checking the global scientific literature. The danger of working this way is not just re-inventing the wheel, which is useful no matter where the idea comes from. The danger is re-inventing a flat tire, which never worked as a wheel in any prior test on the planet. Programs that seem like a good idea based on a System I snap judgment may have already been tested methodically with System II methods. The results may have shown that the idea is not only ineffective but harmful.

Programs that try to "scare straight" at-risk youth are a prime example of the value of checking for international impact evaluations (Petrosino, Turpin Petrosino, and Buehler, 2002). Seven independent tests have shown that, on average, taking pre-adolescent children to tour prisons does not scare them straight; it only increases their participation in crime. Anyone who has the idea of launching that kind of practice is well advised to check the international impact evaluations for the evidence (see, for example, http://www.campbellcollaboration.org/reviews_crime_justice/index.php).

Many other programs that could be just right for local crime conditions can be found by searching international sources. The development of a matrix for matching tests of police strategies to specific crime problems is a recent indication of the growing ease of conducting such searches (see http://gemini.gmu.edu/cebcp/matrix.html).

As the next section shows, however, the most important step in developing locally effective solutions is a systematic analysis of local crime problems. The capacity to conduct such an analysis takes us back to the first principle of evidence-based crime prevention and to our first step in the DRIVER model. Before we take that step, however, we will provide two key elements: definitions and a short user's manual.

**Innovations, Practices, Programs, and Protocols**

This protocol has used several terms almost interchangeably. Yet some distinctions among them are important. An *innovation* can be defined as any practice or program that has not been used in a city or country in recent memory.

A *practice* is almost anything that anyone does with a sincere belief that it may prevent some crimes. A practice does not have to be part of an organized program, but rather something that people can do individually or collectively. A parent spanking a child for punching a sibling may believe doing so will prevent crime by setting limits on violence—even though some evidence suggests that spanking promotes violence by demonstrating the parental endorsement of violence as a solution to conflict (Straus, 1996). A teacher helping students to read may believe it will help prevent crime by promoting literacy and greater employability in a globalizing economy. A nurse reporting a parent for child abuse may be preventing crime by

rescuing the child from an abusive family. A police officer making an arrest for minor domestic violence (such as a slap) may believe it will reduce the risk of further violence.

A *program* can be defined as an organized set of practices that generally involve cooperation among two or more people, usually in the setting of a formal organization. Programs can include the policies of police and criminal justice agencies; schools; and recreational, cultural, and health promotion organizations.

A *protocol* is a set of principles and practices for delivering a program. The term is widely used in medicine to specify the steps to be followed in performing surgery, managing epidemics, delivering babies, and other common procedures. In the hope that the development of citizen security programs can become a more common procedure around the world, this protocol is provided to move such development away from the hunches of System I to the formal analyses of System II.

### How to Use This Protocol

There are two ways to use this protocol. One is by taking one principle of evidence-based prevention at a time and developing innovations to help apply those principles locally. The other way is to follow the entire DRIVER process from start to finish. Unless crime measurement is well established, the full DRIVER process may not be possible. Thus starting with good measures should generally precede the attempt to use them in the full DRIVER process. Once those measures are in place, many DRIVERs can run in parallel.

## The DRIVER Framework

### The **DRIVER** Process: Development Through Evaluation

DRIVER is a protocol for developing, testing, and refining crime prevention practices in an evidence-based process of continuous quality improvement (Deming, 2000). Its acronym denotes the similarity between evidence for preventing crime and for driving an automobile. It is possible, but not advisable, to drive a car without a dashboard of information about speed, fuel, oil, and water, as well as a windshield through which to view one's surroundings. In the same way, it is possible but not advisable to prevent crime without clear and essential information.

### The Six Steps of DRIVER

#### ***D****iagnosis of Specific Problem to be Solved*

DRIVER begins with a diagnosis of the causes and concentrations of the crime patterns of concern. The starting point for diagnosis is *disaggregation,* or breaking big problems into smaller, more manageable parts. Crime is typically categorized as violent, property, drug, or disorder crime. It can also be sliced from other angles as crimes involving people under or over 18 years of age, female or male victims, repeat or first offenders, with or without weapons, by drug-using or non-using offenders, by gangs or not, or by gangs with connections to drug cartels or large-scale organized crime.

There is no one right way to diagnose the crime patterns of any nation or community. The many options available are only more or less useful and appropriate. No matter which option is chosen, some essential tools will increase expert and public understanding of the dimensions and future trajectories of the problem. *Trends* in crime over time in various classifications provide valuable intelligence about the immediate and longer-range future of threats to citizen security. *Patterns* of crime by specific locations, times of day, or modus operandi (such as burglary by distraction, or home invasion robberies to steal car keys and cars) may remain constant even as trends in volume or frequency rise and fall. *Predictions* of individuals who may commit murder

or other serious harm are now increasingly possible with advanced data mining and large databases on criminal histories (Berk et al., 2009).

These tools provide options for a diagnosis that sets *priorities*, making choices about how to define what *targets* to set. Targets in this context means, first, the kind of units of analysis to address, such as repeat crime locations or dangerous offenders in general. Second, choosing targets means listing the specific units within each category, such as particular street corners, young people at risk in high-crime areas, or individual serious offenders recently released from prison.

For maximum efficiency, diagnosis can target the "power few" units within any category of diagnosis: those units with the highest seriousness and volume of crime. Or it might target "low-hanging fruit": units where chances of success are greatest with the least effort. Diagnosis should conclude by setting specific *objectives* as precisely possible. It is far more useful to establish precise dimensions of crime that must be prevented rather than just crime in general, even specific categories of crime. If, for example, banks are being robbed at an increasing rate, reducing bank robberies could be the objective. But if some banks have more customers or employees shot during robberies, then the objective could be reducing injuries and deaths during bank robberies. The two objectives could actually call for different prevention practices.

***R**evision of Practice for Innovative Solutions for Each Problem*

Once each target is identified and selected, a separate course of the DRIVER process can be followed for each target. That course begins with reviewing the current practices and programs that may be directed at each target. This review can be undertaken either for a category of targets (such as banks that are repeatedly robbed each year) or for each individual member of a category (such as the "xxx" Bank branch at 1405 Main Street).

Whatever the target or objective, a review of current practice should be informed by a global and local search for ideas. A global search for impact evaluations on bank security measures, for example, will yield a number of studies that any city can consider (see the DRIVER Protocol Checklist). Similarly, a review of local data may also suggest some pathways for revising prevention practices.

If banks having repeated robberies are less likely to have armed guards than banks that are not being repeatedly robbed, an objective of reducing robberies could be pursued by having police meet with bank managers to suggest they retain armed guards. If, on the other hand, banks with armed guards are less often robbed, but *more* often subject to gunfire exchanges and injuries when robberies do occur, a DRIVER objective of reducing gunshot wounds could lead to recommendations to reduce armed guards.

This example should illustrate how global and local evidence can be linked. Once the DRIVER process exhausts the assessment of current practice, the review of global practice, and the implications of recent crime patterns and trends, it should lead to a decision about what program or policy should be introduced to better achieve the precise objectives set.

### *Implementation of the Revised Practice*

Any effort to introduce change in governmental or community organizations requires an implementation plan, which should include the following:

- a chronological sequence of what will be done
- a clear identification of who will do what, when, and how often
- one or more indicators to be collected for measuring whether the prescribed tasks (outputs) are being performed, and to what extent
- a clear assignment of responsibility for gathering the implementation measures
- a structure for periodic reviews of implementation measures by leadership

### *Value-Added Cost-Benefit Theory of the Innovation*

Before launching an innovative revision of practices or programs, a DRIVER process should also map out the theory of the revision. Building on the implementation plan, the value-added analysis should estimate the extent to which the innovation will be cost-effective.

A mandatory prison sentence for carrying a gun, for example, could lead to a tripling of the prison population—or even a tenfold increase. Analysis of the relevant court data could estimate the cost of the proposed change in terms of prison budgets. It could also estimate the

amount of gun crime that might be prevented solely by the incapacitation of gun carriers. A count of gun crimes committed by people with previous convictions for gun-carrying could serve this purpose adequately and provide a big improvement over sheer guesswork. If the projection estimated that only two gun crimes would be prevented per 100 inmates per year, at $25,000 per year of imprisonment per inmate, the cost would be $2.5 million to prevent two gun crimes—$1.25 million per crime. If only one in four gun crimes resulted in a gunshot wound, the cost per wound prevented would be $5 million.

Even if the value of a gunshot wound prevented cannot be estimated accurately, such a projection would have the benefit of benchmarking the proposal against the cost-benefit ratios of other proposals. It would also lay out a template for repeating the analysis with the actual data from implementing the innovation, if the decision is made to proceed after reviewing the value-added cost-benefit theory.

***E**valuation of the Implementation and Theory of the Innovation*

Once a decision is made to launch an innovation, DRIVER moves into the evaluation of implementation and theory of the innovation. This evaluation can be discussed on an ongoing basis in implementation review sessions. Most importantly, evaluation can provide an early warning to crime prevention leaders that the innovation is not being implemented as planned. If that occurs, the leaders can then decide whether to suspend the innovation, impose mid-course corrections in management or personnel, or continue with clear messages to relevant staff that better implementation is required.

In the case of the Minneapolis hot spots policing experiment (Sherman and Weisburd, 1995), the huge jump in the patrol dosage ratio of target locations to control locations followed a new requirement for police to file daily logs of time spent in their assigned hot spots. The time logs, in turn, were audited against the independent measures of police time in the assigned locations. The result of this constant monitoring from the outset was successful implementation and substantial impact on target crime types.

The theory of any prevention innovation can be spelled out with both measured and unmeasured causal links. In the case of the work of Sherman and Weisburd (1995), the theory was that a greater police presence would increase the general deterrent threat. Part of the theory

depended on having more police presence, which was measured. What was not measured was whether potential offenders actually perceived the threat, or even blinked without thinking that they did not want to commit crimes at locations where police had become so much more visible.

An additional unmeasured dimension of the theory in hot spot policing was the hypothesis that crimes deterred in one location will pop up in other areas. While Weisburd and others have substantially falsified that hypothesis in general, it was virtually impossible to monitor that hypothesis as the innovation progressed. This left the police leadership in some doubt about whether the concentration of patrols in hot spots was lowering crime in the city overall, rather than just moving crime around the corner.

Nonetheless, with a clear objective to reduce specific crimes within the hot spots relative to the same crime types in similar (randomly assigned) hot spots without extra policing, the test of the theory failed to falsify its logic. The innovation was implemented, its theory worked, and the strategy was continued—in the long run, if not the short. Many other cities have adopted the strategy since then, as have countries around the world.

### *Review of New Ways in Which the Practices Could Be Revised and Tested*

Despite the success of any innovation, there is always room to improve on success. In the spirit of continuous improvement, a success should lead to the same question as a failure: how can we do a better job? Asking this question can trigger a return to the beginning of every DRIVER process, including a global review of new research findings, new technologies, or new ideas and programs that have been tried locally or in other places.

This is just what has happened in the two decades since the success of the initial hot spots patrol experiment: many opportunities have been identified for refining and extending the theory of the innovation. Minneapolis itself has subsequently rediagnosed the problem as predicting and preventing emerging hot spots, rather than policing existing hot spots. Jacksonville, Florida, has compared two different policing strategies in high-crime areas, finding that preventive patrol was less effective in the longer term than POPproblem-oriented policing problem-solving (Taylor, Koper, and Woods, 2011). Philadelphia has tested foot patrols in high-violence hot spots, finding reductions in crime, some evidence of displacement, and a net reduction in crime after discounting the benefit for estimated displacement (Ratcliffe et al., 2011).

A DRIVER process that shows no improvement in a crime problem could be the result of a lack of a control group. It is unlikely that a random assignment impact evaluation can accompany every innovation. The Evidence-Based Program Development section discusses quasi-experimental designs for measuring impact, but even these may be too complicated for many organizations to undertake without help from local universities. Thus if the focus is on a crime trend that is not going down, the simplest way forward is to re-analyze the crime that has occurred to detect any changes in the crime pattern. If that re-analysis shows that crime has changed, the question then becomes whether the change fits the theory of the innovation. If not, it may be best to go back to the Diagnosis stage and seek an entirely new strategy. If the crime has changed, it may be best to refine the innovation as developed.

## How *Not* to Design Programs

One good way to develop best practice is to study bad practice. Team coaches often tell athletes how *not* to run or kick a ball; voice coaches tell singers how not to sing a high note. Surgeons are taught how not to make an incision, and architects are told how buildings may fall down if they are designed badly. In many fields, research shows that 10,000 hours of practice is required to do complex tasks well (Gladwell, 2005b; Kahneman, 2011: 238). Yet few people who design citizen security programs have the benefit of any prior practice at all, let alone a coach to show them what not to do.

This section aims to provide a "coach" for people designing security programs who have little or no training for their work. Readers who have already made the mistakes described in this section should be assured that no disrespect for their work is intended. The analysis is offered entirely in the spirit of trying to help everyone design better programs, with better understanding of the DRIVER process.

First, a word on the data: for 10 years at the University of Pennsylvania, DRIVER was the subject of a 15-week, 45-hour course with substantial readings. The course divided top graduate students in government administration and criminology into teams of five. Each team was given a DRIVER focus, such as reducing child abuse or gun violence. Teams were then asked to prepare a 15-minute presentation that covered all six points of the DRIVER process.

Even with all the lectures, readings, and discussions, these teams still made common mistakes that help form the basis for this section.

This section is also based on reviews of hundreds of grant proposals submitted to state and federal agencies in the United States for crime prevention funding and on a few similar proposals submitted for various competitions and funding in Latin American countries. Finally, it also reflects the common mistakes in criminal justice agencies in Europe, Asia, South Asia, Australia, and the Americas that the author has observed since 1970.

Program First, Problem Second

By far the most common mistake is to invent a program and then find a problem to justify it. This mistake is especially likely when the invention is not based on a review of either evidence or theory on the causes and prevention of a particular crime problem. A flash of intuition drives the design of the program, with no other basis for assessing its likely success. This kind of idea can emerge from what British news media call "sofa government" (ideas that come up while people—prime ministers, police chiefs—are sitting on sofas, speaking informally).

**EXAMPLE:** The Drug Abuse Resistance Education (DARE) program. While the program sounds like it is aimed at reducing drug abuse, it was really about sending police into schools to talk face-to-face with pre-adolescent children. The program spread around the United States like wildfire, possibly because the picture of a uniformed police officer in a classroom looked so good on television. It even spread to other countries. Meanwhile, quasi-experimental impact evaluations showed that DARE had no effect on drug abuse by children receiving the program.

The program then became embroiled in discussions about whether the impact evaluations were done properly, or whether the follow-up period was long enough. The subtext of these discussions was the suggestion that the program may not really prevent drug abuse, but that it was very successful in making young children trust and admire police officers. In other words, the commitment to the program for its own sake sparked a process of searching for goals to achieve, rather than starting with a goal and searching for the best program to achieve it. The mayor of Salt Lake City, Utah, reacted to the impact evaluations by cancelling the city's funding for DARE, which led to an attempt by DARE supporters to remove the mayor from office.

Ironically, the issue of trusting police officers rose in theoretical and policy prominence over the same period (National Research Council, 2004). It may well be that DARE is an effective program for enhancing the legitimacy of police institutions, which can be an independent factor in encouraging people to obey the law (Tyler, 1990). Had DARE been designed to improve children's attitudes toward police officers, it may well have been a successful program. The point remains: DRIVER should start with a problem to solve, not with a program to promote.

Program First, Value-Added Theory Second

Another mistake is to design a program or practice intuitively, then struggle to create a theory to explain how and why it should add value to achieving crime prevention goals.

**EXAMPLE:** Arts and recreation programs, such as the Police Athletic League or "midnight basketball," a U.S. proposal rejected by Congress in the 1990s. Like DARE, the sight of young people engaging in such pro-social activities is intuitively appealing. In a System I snap judgment, such activities can receive much support. Yet proponents of these programs are hard pressed to specify how the programs can reduce crime. The claim is that time spent "off the streets" and under adult supervision can keep children out of trouble. That claim depends, however, on how much of the day the children remain under adult supervision and how many days per week they attend the programs. It also depends on the programs not leading to disputes between individuals (or gangs) that provoke violent incidents when children leave the supervised location. Another theory often suggested after this type of program is developed is that recreation improves a child's self-image or self-confidence. Yet there is no evidence that

1) recreation improves a child's self-image, especially in competitive situations in which most participants will either lose to the "power few" champions, or that
2) even if recreation succeeds in making children feel more self-confident that such feelings are a protective factor against crime.

A DRIVER process works better by starting with an established (and well-tested) criminological theory, then designing activities around the theory. An example is control theory (Hirschi, 1969). This theory claims that delinquency is less likely when children have stronger social bonds to conventional morality and society. These bonds can be strengthened by

- *attachment* to families and people who obey the law
- *commitment* to social norms and institutions (school, employment)
- *involvement* in conventional activities
- *belief* that these things are important

This theory implies that children may not actually be better off in child-centered or peer-dominated settings, where conventional morality may be disvalued. It suggests that it may be more important for children to spend time in adult-centric settings, or at least highly structured activities that are directly led by adults under tight control.

Starting with control theory, then, could produce a preference for orchestral concerts over sports, or for choirs over artwork. Any group activity with tight control and an inclusive spirit may foster more attachment, commitment, involvement, and belief than more competitive or isolated individual activities. A DRIVER process could build on this kind of "translation" of theory into practice to choose practices that better fit the theory—and not vice versa.

Multiple Programs without Logical Linkages

There are occasions when a big crime problem provokes a big response. The Inter-American Development Bank, for example, is often asked to support packages of crime prevention programs aimed at improving citizen security. Such packages are logically more likely to succeed, however, if they consist of programs that complement and reinforce each other.

**EXAMPLE:** Suppose a country has a very high homicide rate, with many kidnappings of local people and visiting business executives. Suppose that country could spend or borrow $300 million. How should that country *not* design a package of programs to buy with that funding?

- Decide first how to divide the money among powerful agencies.
- Decide first how to divide funding for the apparent benefit of children, women, and what Argentinian Supreme Court Justice Raúl Zaffaroni calls "heroic victims" for symbolic politics rather than security.
- Pick and choose from a range of proposed programs on a "blink" System I snap judgment about which programs "feel" like they will work.

Using any of these methods of selecting a package of programs is likely to scatter resources and waste money in two ways. One way to waste money is on programs that will not work, one by one. The other way to waste money is through lost opportunity to address profound institutional problems in the nation's crime prevention structures.

If the main problems are homicides and kidnappings, then DRIVER should focus on those two problems. If police failures are highly implicated in both, then the analysis can examine the possible causes of police failures and corruption: poor salaries, low standards of education required at appointment, a sharp dividing line between senior officers and the majority of uniformed police, lack of external accountability, or massive alliances or threats involving major drug cartels.

If the police are a central concern in citizen security, it makes more sense to design a broad package of police reforms than a something-for-everybody menu of disconnected programs. If drug cartels are the main problem, then a package of programs for reducing the harm they cause could become the objective of the DRIVER process. Whether the problems are large or small, focused or comprehensive, DRIVER should always begin with an analysis of where the most harm is coming from. Like a heat-seeking missile, DRIVER should be a harm-seeking process that finds the biggest danger and makes that the first priority.

## Programs without Logic Models

Some leaders oppose impact evaluations because they do not want to subject the logic model of their programs to careful scrutiny. That is exactly why a requirement for impact evaluation is so important for every innovation. A logic model, or theory of how the program should work, is

essential to spell out the connection between the innovation and its objective. Yet many crime prevention innovations fail to identify their objectives, let alone a logic model for accomplishing them.

**EXAMPLE:** A new police academy is to be established. The students, the content of the teaching, the purpose of the academy—these are all left unspecified in a request for funding. One performance indicator is provided: how many criminal cases are solved by the graduates of the academy. Yet there is no discussion of how criminal cases are investigated or whether different kinds of cases may go to graduates of the new academy than of the old academy. It is not even clear that cases are solved by lone investigators, rather than teams that may combine graduates of the old and new academies.

A new police academy could be a very good idea as part of a DRIVER process to create a more effective, fairer, and less corruptible police institution. There does not even need to be a specific crime reduction objective if the process was aimed at better policing per se. The improvement of citizen security does not require a direct connection between each innovation and a specific type of criminal event. Rather, it requires great clarity about cause and effect. A theory that a new police academy will improve police integrity and competence is a clear logic model. The model can have multiple objectives or indicators. But the big idea is elegantly clear, no matter how complex the details. Such transparency is the opposite of the opaque, indiscernible purpose of just establishing a new police academy without a logic model.

Programs without Local Data

It is a huge mistake to design an innovation without first diagnosing the problem locally. Many security programs, sadly, are designed without access to local crime data. Anecdotes and newspaper stories may be all that developers can reference. Yet if that is the situation, it is not an obstacle to solving an important problem. It *is* the most important problem.

**EXAMPLE:** A school proposes to develop an after-school program for children in a high-crime area to reduce their risk of involvement in delinquency. The school has no data on how many of its students, in each age group, have already been charged with delinquency. It has no capacity to

use those data to target its efforts, or even to ensure that delinquent children are not concentrated in certain groups. Worst of all, it has no capacity to measure delinquency among children who have completed the program or are still in it. There is, then, no way for this apparently good idea to be tied to crime prevention.

One could argue that a school's purpose is not to make crime data more reliable and accessible. One may just as well argue that schools have nothing to do with crime. But the need for reliable and useable crime data is so great that all institutions concerned with crime can be asked to start with the data first. On a firm foundation of good crime data, even with careful respect for privacy rights, an unlimited opportunity for crime prevention can be built.

Programs without Prior Evidence on Problems

An issue related to the lack of local data is a lack of global data. Crime is changing rapidly worldwide as a result of technology and communication. If the crime problem of interest has never occurred in one locality, it may be possible to learn something about it from other localities where it has occurred. If the problem has never occurred before anywhere, the challenge is greater. The best response may be to keep looking for evidence of similar patterns.

**EXAMPLE:** "Flash mob" riots came to England in 2011. They had never been seen before in that form: hundreds or thousands of people being told to assemble at a certain location where they could loot and burn. The instructions came by text message, Twitter, and other instant media on personal digital communication devices. Police were not prepared to have mobs move around London faster than police themselves could arrive in sufficient numbers. Parliament debated the matter and considered using water cannons and rubber bullets. Yet there was little evidence presented about who was involved, or about what the dynamics of such public order problems might be more generally.

In the months after the August riots, more evidence was gathered on who was arrested and what their prior crime records were. Yet little evidence has been compiled on flash mob behavior, which has apparently been used for several years to organize parties rather than riots. Crafting prevention plans for such behavior is difficult, but can only benefit from more global research on how such behavior develops.

Programs without Prior Evidence on *Solutions*

In general, it is unwise to implement programs without any evidence of the likely effectiveness of the solution they would impose on a certain problem. A solution may not have been tested in a randomized controlled experiment, but there may still be some evidence on the solution in the global evaluation literature. Ignoring that evidence can lead to adopting ineffective or even harmful programs.

**EXAMPLE:** Second responders follow up on domestic violence. This solution to domestic violence is a supplement to police responses to a domestic disturbance in progress. When police visit a home in an emergency, they usually have no special skills to deal with a conflict. Several programs in the United States have dispatched a social worker or counselor with police within a week or two after an emergency call for police at that address. These programs have been assessed with several strong randomized experiments—one of them in Miami with a largely Latino population. The results show that such visits only increase calls to police, with no evidence of reduced injury or any other benefit. Any jurisdiction that adopted such a program would be ignoring good evidence that money spent on it would be wasteful at best. At worst, the program could be causing more violence in the home.

"Black Box" Programs: No Measures of Key Implementation Steps

Perhaps the worst way to design a program is to have no direct measure of whether the program is even delivered. Even with a good impact evaluation, it is hard to know how to interpret the data on *outcomes* without data on *outputs*. This leaves the funders of the program unclear about whether the program did not work or whether it was ever even delivered. Did the pill fail to cure the disease—or did the nurse fail to give the patient the pill? (Or did the nurse give the pill to the patient, but the patient did not take it?).

**EXAMPLE:** A cognitive behavioral therapy (CBT) program is introduced in a prison. An impact evaluation examines the recidivism rate of offenders coming out of the prison. The

recidivism rate declines in the year after the CBT program is introduced relative to the year before. But no data are provided on how many inmates actually began CBT, let alone completed the 12 one-hour sessions recommended in the evidence review of previous evaluations. Whether 100 percent of the inmates or 5 percent received CBT is unknown. No DRIVER process can make any sense of such a "successful" program.

"No Outcomes" Programs: No Measures of Problem Impact

The most common mistake in designing programs may be not reviewing any outcome data at all on a program in a way that ties the outcomes specifically to that program. Even if there are good output data on how many people have attended how many sessions of a program over a specific period, there may be no information on how many crimes those people committed later—or how many crimes were committed against them. DRIVER cannot reach any conclusions about whether to continue, stop, or refine the program.

**EXAMPLE:** A social agency hires "detached street workers" 25 to 35 years of age to become friendly with teenage gang members. The street workers each spend long hours with a group of 10 to 30 teenagers who live in a certain area. They keep good records on who the teens meet with, what pro-social activities they organize, what places they visit, and even what (legal) businesses they may establish. No data on the criminal activities of the teenagers, however, are available to the social agency, nor do the street workers keep track of arrests they may hear about from the teenagers directly. Without such information, no DRIVER process can reach completion.

"No Comparison" Programs: No Measure of Value Added

The most frustrating mistake a program can make is to do all but one thing right. A DRIVER process could select an evidence-based program with a good theory, but in the absence of a comparison to a similar group not using the program (or using a different program), it could reach entirely misleading conclusions.

**EXAMPLE:** A program for juvenile delinquents could measure its intended outputs, such as meetings between counselors and children. It could measure an outcome before and after the program, such as the frequency with which children are arrested. It could do everything except estimate the value added: the true difference the program made in any trends in crime. It may also suggest that the program failed when it really succeeded. Because the arrest rates of the children in the program went up, the assumption might be that they were going to rise anyway, and the program failed to stop it. Yet with a valid comparison group, the evidence could show that arrests rose by much less than among a group of similar children who were not enrolled in the program.

A comparison may be hard to find. But even if one is unavailable, benchmarks can be used to project a virtual comparison. It is even possible to use the children enrolled in the program as their own controls, by comparing their actual performance against their predicted performance (Sherman, 1997. No matter how it is computed, a fair means of comparing the served population to a "counterfactual" case without the program is essential to the DRIVER process.

## Evidence-Based Program Development

### What Is Evidence?

Evidence consists of systematically collected observations in support of a conclusion. The use of such observations in designing citizen security has been relatively thin compared with best practices in business or health-care planning. These and other fields require high levels of System II evidence to support large investments. Citizen security policy has been far more oriented to common sense, System I theories, and traditional practices and customs. The difference between evidence-based development of citizen security programs and previous practice is largely a massive increase in the extent to which conclusions need to be support by robust evidence, evidence that can withstand alternative interpretations and theories that can undermine a conclusion.

The two key conclusions needed in citizen security programs require evidence on *predictions* and *prevention*. Both are major arenas for contests between Systems I and II. Both are also essential for DRIVER processes. Both can begin with systematic efforts to describe patterns and trends in crimes and the populations of criminals.

## Patterns and Trends

A pattern is any repeated distribution of events, such as the sun rising each morning and setting each night. Crime patterns come in many varieties, from modus operandi (murders in barfights with knives) to motives ("honor" killing of Muslim women who break Sharia law) to demographics (bank robberies by elderly men in wheelchairs). A trend is a kind of pattern that consists of changing frequency of events or rates over time—up or down. Patterns and trends in crime can lead to selecting problems to address or to disaggregating the key targets within broader problems. The fact that half of all murders this year have been committed in three neighborhoods that had no murders last year, for example, is both a pattern and a trend. It can open up many further analyses and ideas for ways to prevent further murders. It can also open the door to prediction.

## Predictions

Predictions are forecasts that something will or will not happen, such as rain or no rain (a binary outcome), or the degree to which something will happen, such as how much rain will fall (a continuous outcome). The examples could just as easily be whether there will be a murder this weekend in this city, or how many murders will occur. They could also be whether this offender will commit another crime on release from prison, or how many crimes that particular offender will commit.

System I predictions like these are made on the basis of experience and hunches, or even detailed case studies. System II predictions are made by statistical analysis of very large databases, which are essential for making accurate predictions of rare events (Meehl, 1954). Five decades of contest between the "clinical" System I approach and the "statistical" System II approach has resulted in consistent victories of statistical over clinical prediction (Kahneman,

2011). Nonetheless, many criminal justice agencies rely on clinical predictions, while few have adopted statistical methods.

Statistical prediction is not only more accurate but it can operate at multiple levels. Predicting a neighborhood in which a murder may occur can lead to reviewing individual offenders who might be most likely to commit the next murder. Statistical analyses of both neighborhood and individual data open the door to making multi-level forecasts with clear implications for prevention strategies.

## Prevention: The Maryland Scale of Scientific Methods

Once a prediction is made that a unit is at high risk of a crime problem, prevention programs can attempt to intervene to reduce the risk. Whether the intervention is effective is a completely separate question from whether the prediction was correct. Yet the only way to know whether the intervention worked is to compare units exposed to the intervention to units that were not. (This principle is number 7 of the 10 principles listed in the executive summary.)

When a comparison is made, it needs to be as fair (unbiased) a comparison as possible. Ideally, the only difference, on average, between units receiving the intervention and those not should be just that: whether they received the intervention. The closest approximation to that ideal is generally found when comparisons are made by random assignment experiments, also known as randomized controlled trials (RCTs). Prevention efforts evaluated in this way are placed at the top of the Maryland Scale of Scientific Methods (Sherman, 1997), a 5-level scale that ranks the level of "internal validity" of the most commonly used crime prevention evaluation designs. The scale is as follows:

> **Level 1.** Correlation between a crime prevention program and a measure of crime or crime risk factors
>
> **Level 2.** Temporal sequence between the program and the crime or risk outcome clearly observed, or a comparison group present without demonstrated comparability to the treatment group

**Level 3.** Comparison between two or more units of analysis, one with and one without the program

**Level 4.** Comparison between multiple units with and without the program, controlling for other factors, or a nonequivalent comparison group with only minor differences evident

**Level 5.** Random assignment and analysis of comparable units to the program and comparison groups

In general, the best evidence available would use a research design with the highest possible score in the range of 1 to 5. The reason is that the higher the level on the scale, the greater clarity there is about whether the prevention program caused any difference between the targeted and comparison units, or whether the appearance of such causation was actually misleading—because some other, uncontrolled factor was the true cause.

When reviewing previous tests of any program, it is more certain that the result of a higher-ranked design is correct. However, certainty is not the only consideration. It is arguably more important to pick a program with a big effect, or a large amount of prevention per amount of money invested, than to pick one that is very clearly causing a weak effect (a low return on investment). As the Campbell Collaboration says, "effect size matters." Programs with big effects in prior tests might be the best choice, even if they may be less certain about those effects. Even, for example, if the results of previous tests have not been statistically significant, as long as the results have all been in the right direction and are large in magnitude.

In that context, the words "statistical significance" should be read simply as "not clear," rather than "weak." There are many strong effects that may be of borderline clarity, in which case the effect size should be the primary indicator of whether to select the program for local use. All this is subject to further considerations, as described below.

## Previous Tests of Similar Programs

Choosing a prevention program is, in effect, a prediction that the program will work. The best basis for such a prediction is previous tests of the same or similar programs. Once a DRIVER

process has examined local data, it is very important to search the Internet for evidence of previous tests of any program that has been considered for introduction into a jurisdiction.

The best evidence would be a systematic review of repeated random assignment experiments testing the same program or practice against the same comparison groups. Repeat experiments provide more *external* validity, also called "generalizability," defined as confidence that what worked there will work here.

If no such review is available, a single randomized experiment may provide better evidence than repeated quasi-experiments. One good experiment can be more persuasive than several less conclusive quasi-experiments because it has greater *internal* validity, having ruled out more competing explanations in that one study as threats to its conclusion.

If no randomized trials can be found, repeated quasi-experiments are the next best evidence. They are most encouraging when all or most of their results go in the same direction, even if the effects are not statistically significant. The odds of repeated results in the same direction (such as reducing crime) occurring by chance are quite low. That means that a *pattern* of results can be statistically significant even when the individual study findings are not.

Previous Tests of Solutions for Similar Problems

If no tests of solutions to a particular problem can be found, evidence-based program development can always rely on the next closest evidence. A test of a solution for a similar problem with a different population, or a different stage of the criminal justice process, can be reasonably generalized to the problem at hand.

**EXAMPLE:** There is as yet no randomized experiment on making specific threats to arrest gang members for minor law violations—a strategy often called "pulling levers." There is, however, an experiment in making specific threats to probationers who are regularly tested for using drugs on probation. The probation experiment (called Project HOPE) evaluated the use of immediate jail for a single night the first time a probationer failed a drug test, two nights for the second failure, three nights for the third, and so on. The overall effect of random assignment to that treatment was far fewer arrests than among similar probationers not assigned to that treatment (Hawken and Kleiman, 2009). That experiment constitutes direct evidence on the problem of

crime by probationers. It may also, however, constitute indirect evidence on the effects of threats of immediate arrest of gang members if any member of their gang kills somebody (Weisburd, Sherman, and Maher, 1993).

## The Iterative Process of Analysis and Design

Evidence-based program development is *iterative:* a back-and-forth progression between evidence and conclusions, new data and new actions. The DRIVER model is entirely contingent on the evidence as we know it today; conclusions can change with new evidence tomorrow. The scientific view of evidence is that it is always incomplete but can always be expanded. That view stands in sharp contrast to a legal or ideological position that is entrenched and unchanging as a matter of principle. The only room for such intransigence in evidence-based program development is for human rights and bright-line ethical positions. All other conclusions about crime prevention practice should stay ready for modification as a result of new evidence. Otherwise, we might still think the earth is flat.

## Connecting Security to Economic Development

The iterative DRIVER process for enhancing citizen security is no different from the evidence-based process of economic development. Both rely on theories of what works, which are more or less imperfectly tested by evidence. Both rely on systematic descriptions of problems, such as who is unemployed (or victimized by crime), where, and why. Both may also be solved by similar programs.

Scientific knowledge about the effects of economic development on crime is still thin. There are many surprising findings of crime growing with the economy and of employment programs failing to prevent crime. Yet there is also powerful evidence of crime falling with income inequality, as well as with inequalities of wealth. Homicide in particular is related to the Gini Index of income inequality; so is the rate of killings of citizens by police (Sherman and Langworthy, 1979). It may be that employment rising with inequality is little help against crime, but that long-term strength of employment could promote falling inequality and thus reductions in violent crime.

A more important reason to connect security to economic development may be the harm crime causes to the economy. The evidence that in the United States, the costs of crime constitute 17 percent of gross domestic product, including direct costs of crime as well as funds spent on security and justice, is a startling reminder of the role of crime as a hidden tax on economic growth (Ludwig, 2006). Taking a long view of economic growth may require that any nation make itself safe for business, especially in a globalizing economy.

## The DRIVER Protocol: A Checklist

This checklist provides a step-by-step guide to developing evidence-based programs for citizen security. It is presented as a series of questions to be answered very briefly but precisely. It can be used for several purposes, including recording decisions made over a series of meetings, or milestones in agreements on how to shape a proposed program. It can also support requests for program funding or by funding agencies to assess the strength of a case for investment. Its best use could be by organizations that both develop and deliver programs.

There is good evidence that checklists can improve performance in a wide range of settings, including the prevention of airplane crashes, surgical errors, and nuclear accidents (Gawande, 2009). The effect of using this checklist may be to find more and better evidence that can lead to cost-effective practices and programs for crime prevention. As Malcolm Gladwell observed in his review of Atul Gawande's book *The Checklist Manifesto*,

> Gawande begins by making a distinction between errors of ignorance (mistakes we make because we don't know enough), and errors of ineptitude (mistakes we made because we don't make proper use of what we know). Failure in the modern world, he writes, is really about the second of these errors…

The checklist, like DRIVER itself, follows a classic problem-solving action-research cycle, first proposed by psychologist Kurt Lewin (1958). At the end of each cycle, it may be helpful to start a new checklist. For problems of great urgency, it could be useful to refresh a DRIVER process every year. That may help ensure that the process stays up to date with both global and local evidence.

### Step 1: Diagnosis of Specific Problem to be Solved: Concentrations and Causes

1.1 What is the specific problem to be solved? (e.g., inter-gang homicides and attempted homicides)

1.2 What categories of the problem are most relevant to any intervention? (e.g., specific gangs)

1.3 Which specific categories should be given the highest priority and why? (e.g., the three most deadly gangs)

1.4 What measurable data are available for analysis? (e.g., police homicide and assault files)

1.5 If no directly measureable data are available for the problem, what data come closest to measuring the problem defined? (e.g., court records of homicide and attempted homicide convictions)

1.6 Power few analysis: What percentage of the problem is concentrated among what percentage of the following units?

- __ percent of known offenders committed __ percent of the crime.
- __ percent of known victims suffered __ percent of the crime.
- __ percent of crimes occurred between these hours of the day: __:00 and __:00.
- __ percent of crimes occurred on these days of the week: _____, _____, and _____.
- __ percent of crimes occurred at __ percent of the (street addresses) (small areas).

1.7 Demographics of known or convicted offenders:

percent male =
percent under 18 =
percent under 25 =
percent over 40 =
percent with previous convictions =
percent with previous arrests =
percent with previous imprisonment =
percent from a major racial group =
percent residing in high-poverty areas =
percent wealthy =

1.8 Demographics of victims of _____:

percent male =

percent under 18 =

percent under 25 =

percent over 40 =

percent with previous convictions =

percent with previous arrests =

percent with previous imprisonment =

percent from a major racial group =

percent residing in high-poverty areas =

percent wealthy =

1.9 Motives: Describe up to three most common motives, such as gang rivalry:

1_____

2_____

3_____

1.10 Method or technology (means) (e.g., weapons, fraud, corruption of police):

1.11 What patterns of response are currently provided by any of the following?

- Policing
- Prosecution
- Bail
- Convictions
- Sentences
- Correctional programs

1.12 What programs of prevention, if any, are currently in operation?

- Government
- Family
- Schools

- Legitimate labor market participation
- Security systems at places (such as homes of victims)
- Security systems in situations (such as entering taverns)

1.13 What does the global literature say about the macro-causation of the problem? (e.g., income inequality, weak state institutions, weapon density, concentrated poverty areas)

1.14 What does the global literature say about the micro-causation of the problem? (e.g., competition in drug trafficking, extortion from businesses, low prices for alcohol, family networks of criminal offenders)

1.15 What measured trends are available for the problem and what do they show?

- 5–10 year national trend (e.g., the crime is up 37 percent)
- 5-year local trend (e.g., the crime is up 44 percent)

1.16 What is the most likely reason for the recent trends?

1.16a What is the evidence for that theory, if any?

1.17 What maps have shown the geographic concentrations of the problem? (summarize and attach)

1.18 Describe the chain of causation: Name all the links in the chain of causation that contribute simultaneously, or in some sequence, to the repeated occurrence of these crimes:
(e.g., widespread availability of firearms). (Create as many lines as needed.)

1 _____
2 _____
3_____
4_____
5_____

1.19 Breaking the chain: Which of these causal links would be most likely to reduce the problem if its connection could be broken?

1.20 The weakest link: Which of these causal links would be easiest to break, even if doing so did not have the greatest impact?

1.20a Why would breaking this link be easier than breaking other links in the chain of causation?

1.21 What target is the focus of the initiative you propose to develop? (e.g., reducing the carrying of guns in public)

1.22 List any criminal justice, governmental, or private organizations to which you have presented this diagnosis, and summarize their responses:

**Step 2: Review of Possible Programs for Solving Each Problem**

2.1 Searching English-language Internet sources

2.1a What keywords did you enter?

2.1b Check each database you searched.

www.crimesolutions.gov __

www.campbellcollaboration.org __

www.Home-Office.gov.uk __

www.ncjrs.gov __

www.Google.com __

Google Scholar at Google.com __

2.2. Searching Spanish language data sources

2.2a What keywords did you enter?

2.2b Check each database you searched.

www.Google.com __

(list others) _____

2.3 Selecting relevant program descriptions: Describe one or more programs you found in your web searches that appeared to be most promising for use in your city:

2.4 Selecting relevant impact evaluations: Which of the promising programs you found on the web had undergone impact evaluations that met a minimum standard of Level 3 on the Maryland Scale of Scientific Methods? (see Prevention: The Maryland Scale of Scientific Methods, above)

2.5 Systematic reviews: Did any of the programs that seemed promising have repeated impact evaluations summarized in a systematic review or meta-analysis? If so, summarize the results:

2.6 Fitting programs tested elsewhere to local conditions: Which programs, if any, did you find from other places that seem appropriate (in theory) for local use?

2.7 Which program did you find that could be modified for local use, and how would you modify it?

2.8 Inventing a new program: If no program evaluated or described elsewhere seems appropriate, describe a new initiative you would like to implement that appears not to have been previously attempted elsewhere:

2.9 Developing a logic model: Describe, in theory, the chain of causation between the program you propose to introduce and the measureable result you hope to achieve (e.g., police checkpoints to reduce gun carrying and gun homicides):

> 1. Identify small areas of high gun crime → 2. Police conduct random searches for guns in cars or carried by people entering area → 3. People carry fewer guns in those areas → 4. Fewer shootings occur in those areas → 5. People die of gunshot wounds in those areas

1._____ → 2._____ → 3._____ → 4._____ → 5._____

2.10 Testing the new idea against thought experiments: Describe the result of any "thought experiments" you may have conducted with focus groups that have talked through the logic model above.

2.11 A provisional decision: Describe the main elements of the program to be implemented, if the program is found to provide a cost-effective value to reducing the crime problem:

**Step 3: Implementation Planning for a Proposed Program**

3.1 From logic model to action plan: Summarize, in chronological sequence, the steps that must be taken to implement the proposed initiative, using the chart provided in Table 1 (which contains an example that should be deleted to make room for your proposal); add columns and rows as needed:

**Table 1. Implementation Planning**

| Task | Responsible party | Startup | Quarter 1 | Quarter 2 | Quarter 3 | Quarter 4 |
|---|---|---|---|---|---|---|
| Map gun crime areas | NGO "Stop Gun Violence" | X | X | X | X | X |
| Select police | Police liaison | X | | | | |
| Train police | NGO and Police Academy | X | | | | |
| Assign police to checkpoints | Police crime analysis and NGO | X | | | | |
| Purchase bullet-proof vests for police | NGO | X | | | | |
| Conduct evening searches | Police | | X | X | X | X |
| Conduct early morning searches | Police | | X | X | X | X |
| Conduct daytime searches | Police | | X | X | X | X |

| | | | | | | |
|---|---|---|---|---|---|---|
| Assess progress and problems in monthly implementation review | NGO and Police | | X | X | X | X |

3.2 From action plan to planned measures of actions: List all the measures of actions (outputs) that will be reviewed each month (or more often):

3.3 From measures of actions to managing actions: Describe what actions should be taken if the measures show that the program is not being implemented as planned**:**

3.4 Anticipating obstacles: What obstacles to implementation are expected, and what is planned for overcoming them?

3.5 Designing out obstacles: What anticipated obstacles could be avoided by designing the program differently?

**Step 4. Value-Added Cost-Benefit Theory of the Proposed Program**

4.1 How many crimes happened last year in the target population? Report the best estimate of how many crimes in total happened in the area or among the people to which the proposed program will be addressed:

Number of Crimes _____ Source of Data _____

4.2 Projecting a reasonable effect size for the proposed program: By what percent do you expect the program to reduce the volume of crime in the targeted population in one year?

__ percent

4.3 Estimating total crimes prevented: Applying the percent reduction to the total crimes reported last year, what is the total number of crimes that the reduction would prevent in the first year of the program?

__crimes of (list offense types) _____ prevented

4.4 Projecting the value of the benefit of that effect size: Using any cost data available, project the value of preventing one crime (see Cohen, 1979); Cook and Ludwig (2002); Home Office, 2011. Include any information on the cost per crime of the following, and sum the costs across all dimensions:

| | |
|---|---|
| Medical treatment of victim | ____ |
| Lost days of work | ____ |
| Crime victim compensation | ____ |
| Investigation | ____ |
| Prosecution | ____ |
| Imprisonment | ____ |
| Other costs (list) | ____ |
| Total cost per crime | ____ |

4.5 Projecting the one-year cost of the proposed program: List all costs of the program, both from external and existing operating costs, and sum the cost for one year:

| Cost Factor | Amount |
|---|---|
| _____ | ____ |
| _____ | ____ |
| _____ | ____ |
| _____ | ____ |
| Other | ____ |
| Total cost | ____ |

4.6 Projecting the cost per crime prevented by the proposed program: Divide the total cost of the program by the number of crimes prevented.

One-year program cost = _____ / crimes prevented = _____

4.7 Cost-benefit ratio: Divide the cost of the program per crime prevented by the cost of each crime, and express as a ratio of crime cost to program cost:

Cost of program per crime / cost of each crime to society = ____to ___

4.8 Is the proposal cost-effective?

Enter "yes" if the cost per crime exceeds the cost per crime prevented.

Enter "no" if the cost per crime is less than the cost per crime prevented.

4.9 Next step.

4.9a If the proposal *would not* be cost effective, conclude this DRIVER process and open a new process.

4.9b If the proposal *would* be cost-effective, continue to Steps 5 and 6.

**Step 5: Evaluation Planning for the Proposed Program**

5.1 Will the evaluation be a randomized controlled trial? If the proposed evaluation can be conducted as an RCT, go to http://www.crim.cam.ac.uk/research/experiments/beck-crimport1.0.pdf for an example of an experimental protocol and the form to follow in preparing a protocol for the proposed project. Completing the CRIMPORT protocol is a complete substitute for this part of the checklist. If an RCT will not be used, complete the rest of this section.

5.2 What are the target units of analysis? (check one)

| | |
|---|---|
| Cities | ___ |
| Communities or neighborhoods | ___ |
| Street-blocks | ___ |
| Street-corners | ___ |
| People (offenders, victims, officials) | ___ |

5.3 How will cases be selected?

5.4 How will cases be chosen to receive an innovation?

5.5 How will comparison units not receiving an innovation be chosen?

5.6 How will the evaluation measure resources invested?

5.7 How will implementation (outputs) be measured for both comparison and target units?

5.8 How will outcomes be measured for both targeted units and the comparisons?

5.9 How will outcomes for the target units be compared with outcomes for comparisons?

You are now ready to implement the proposed program.

---

5.10 Results of analysis: Based on the evidence from the initial test period, what results can be reported using the analysis plan described above?

Implementation ratio of outputs: Target versus comparisons =
Implementation cost ratio of outputs: Target versus comparisons =
Impact ratio of outcomes: Target versus counterfactual =
What is the direction of the effect (target did better or worse?)
What is the statistical probability that the effect occurred by chance? (P = __)

**Step 6. Review of the Program at End of Test Period**

This step is intended to be implemented at the end of an initial test period. Some innovations can be tested quickly, while others may take longer. The timing depends in part on the volume of crimes and the nature of the practice or program. The step can be implemented anytime there is concern that the program may not be viable or that it requires a mid-course correction.

6.1 Implementation: Has the program been implemented largely as planned?

6.1a If not, what elements are not being implemented?

6.1b What are the main reasons for the failures of implementation?

6.1c Is the implementation failure correctable? If so, how?

6.2 Costs: Are the costs of the program within budget?

6.2a If not, what costs are running over budget?

6.2b What are the main reasons for cost overruns?

6.2c Is the cost overrun correctable? If so, how?

6.3 Staffing: Does the program have adequate staff?

6.3a If not, which staff need to be added?

6.3b What are the main reasons the staff are unavailable?

6.3c Is the short staffing correctable? If so, how?

6.4 Impact: Is the program having its intended impact to any degree?

6.4a If not, why not?

6.4b. Is the lack of impact correctable? If so, how?

6.5 Cost-effectiveness: Is the impact large enough to be cost-effective?

6.5a If not, what is the main reason the impact is not large enough?

6.5b What is the main cause of the underperformance of the program?

6.5c Is the inadequate impact correctable? If so, how?

6.6 Next step: What will be done next? (check one)

6.6a The program will be continued in its present form. ___

6.6b The program will be discontinued and a new DRIVER process begun. ___

6.6c The program will be modified and continued. ___

(describe modifications)

## References


Ariel, Barak, Lawrence Sherman, and Jordi Vila . 2012. "Random Assignment without Tears: How to Stop Worrying and Love the Cambridge Randomizer." *Journal of Experimental Criminology* 8(2): 193–208.

Beccaria, Cesare. 1764. *On Crimes and Punishment*. London.

Bentham, Jeremy. 1781. *An Introduction to the Principles of Morals and Legislation.* http://www.utilitarianism.com/jeremy-bentham/index.html

Berk, Richard A., Lawrence Sherman, Geoffrey Barnes, et al. 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 (part 1): 191–211.

Braga, Anthony A., and Brenda J. Bond. 2008. "Policing Crime and Disorder Hot Spots: A Randomized, Controlled Trial." *Criminology* 46: 577–607.

Bratton, William. 1998. *The Turnaround: How America's Top Cop Reversed the Crime Epidemic*. New York: Random House.

Campbell, Donald, and Julian Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Cohen, Laurence, and Marcus Felson. 1979. "Social Change and Crime Rates: A Routine Activity Approach." *American Sociological Review* 44: 588–608.

Cook, Philip and Jeins Ludwig. 2002, *Gun Violence: The Real Costs. Oxford University Press.*

Di Tella, Rafael, and Ernesto Schargrodsky. 2010. *Criminal Recidivism after Prison and Electronic Monitoring*. Cambridge: National Bureau of Economic Research. http://www.pazciudadana.cl/docs/ext_20101129105830.pdf.

Gawande, Atul. 2010. *The Checklist Manifesto*. New York: Henry Holt.

Gladwell, Malcolm. 2005a. *Blink: The Power of Thinking without Thinking*. Boston: Little, Brown.

———. 2005b. *Outliers*. Boston: Little, Brown.

Goldstein, Herman. 1990. *Problem-Oriented Policing*. New York: McGraw-Hill.

Hawken, Angela, and Mark Kleiman. 2009. *Managing Drug Involved Offenders with Swift and Certain Sanctions: Evaluating Hawaii's HOPE*. Washington, DC: National Institute of Justice. https://www.ncjrs.gov/pdffiles1/nij/grants/229023.pdf.

Hirschi, Travis. 1969. *Causes of Delinquency*. Berkeley, CA: University of California Press.

Home Office. 2011 . Revisions made to the multipliers and unit costs of crime used in the Integrated Offender Management Value for Money Toolkit. http://www.homeoffice.gov.uk/publications/crime/reducing-reoffending/IOM-phase2-costs-multipliers

Jackson, Mark T. 2011. MSt Thesis, Institute of Criminology, University of Cambridge, Police Executive Programme.

Johnson, Stephen. 2010. *Where Good Ideas Come From: The Natural History of Innovation*. New York: Riverhead Books.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Lewin, Kurt. 1958. *Group Decision and Social Change*. New York: Holt, Rinehart and Winston.

Ludwig, Jens. 2006. Testimony before the US Senate, Committee on the Judiciary, September 19. Posted at http://www.judiciary.senate.gov/hearings/testimony.cfm?id=e655f9e2809e5476862f735da11bb8b4&wit_id=e655f9e2809e5476862f735da11bb8b4-1-3.

McCord, Joan. 1978. "A Thirty-Year Follow-Up of Treatment Effects." *American Psychologist* 33: 284–289.

———. 2003. "Cures That Harm: Unanticipated Outcomes of Crime Prevention Programs." *Annals of the American Academy of Political and Social Science* 587: 16–30.

Meehl, Paul. 1954. *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis: University of Minnesota Press.

Millenson, Michael. 1997. *Demanding Medical Excellence: Doctors and Accountability in the Information Age*. Chicago: University of Chicago Press.

National Research Council. 2004. *Fairness and Effectiveness in Policing: The Evidence*. Washington, DC: National Academies Press.

Olds, David, John Eckenrode, Charles R. Henderson, Jr, et al. 1997. "Long-Term Effects of Home Visitation on Maternal Life Course and Child Abuse and Neglect: Fifteen-Year Follow-up of a Randomized Trial." *JAMA* 278(8):637–643.

Palmer, Ian, and Andrea Caddick. 2010. Paper presented at the Society of Evidence-Based Policing. Manchester, United Kingdom.

Petrosino, Anthony, Carolyn Turpin Petrosino, and John Buehler. 2002. *"Scared Straight" and Other Juvenile Awareness Programs for Preventing Juvenile Delinquency*. Oslo: Campbell Collective. www.campbellcollaboration.org/lib/download/13/.

Popper, Karl. 1959. *The Logic of Scientific Discovery*. New York: Basic Books.

Ratcliffe, Jerry, Travis Taniguchi, Elizabeth R. Groff, and Jennifer D. Wood. 2011. "The Philadelphia Foot Patrol Experiment: A Randomized Controlled Trial of Police Patrol Effectiveness in Violent Crime Hotspots." *Criminology* 49: 795–831.

Renzema, Marc, and Evan Mayo-Wilson. 2005. "Can Electronic Monitoring Reduce Crime for Moderate to High-Risk Offenders?" *Journal of Experimental Criminology* 1: 215–237.

Sherman, Lawrence. 1992. *Policing Domestic Violence: Experiments and Dilemmas*. New York: Free Press.

———. 1997. "Thinking about Crime Prevention." In Lawrence Sherman, et al., *Preventing Crime: What Works, What Doesn't, What's Promising*. Report to the U.S. Congress. Washington, DC: U.S. Department of Justice. https://www.ncjrs.gov/works/chapter2.htm.

———. 2007. "The Power Few Hypothesis: Experimental Criminology and the Reduction of Harm." *Journal of Experimental Criminology* 3(4): 52.

———. 2010. "Less Prison, More Police and Less Crime: How Criminology Can Save the States from Bankruptcy." Lecture Presented to the National Institute of Justice, Washington, DC. http://www.nij.gov/multimedia/presenter/presenter-sherman/.

———. 2011a. "Criminology As Invention." In Mary Bosworth and Carolyn Hoyle, *What Is Criminology?* Oxford: Oxford University Press.

———, and Dennis P. Rogan. 1995. "Effects of Gun Seizures on Gun Violence: Hot Spot Patrols in Kansas City." Justice Quarterly 12(4).

———, and Robert Langworthy. 1979. "Measuring Homicide by Police Officers." *Journal of Criminal Law and Criminology* 70(4): 546–560.

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

———, Patrick R. Gartin, and Michael E. Buerger. 1989. "Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place." *Criminology* 27: 27–55.

———, and David Weisburd. 1995. "General Deterrent Effects of Police Patrol in Crime Hot Spots: A Randomized, Controlled Trial." *Justice Quarterly* 12(4): 635–648.

———, Heather Strang, and Barak Ariel. 2011. "Experiments without Grants: How to Conduct Low-Cost Randomized Trials." Paper presented to the American Society of Criminology, Division of Experimental Criminology.

Soares, Rodrigo R., and Igor Viveiros. 2010. "Organization and Information in the Fight against Crime: An Evaluation of the Integration of Police Forces in the State of Minas Gerais, Brazil." Bonn: Institut zur Zukunft der Arbeit [Institute for the Study of Labor]. Discussion paper 5079.

Straus, Murray. 1996. “Spanking and the Making of a Violent Society.” *Pediatrics* 98: 837–842.

Taylor, Bruce, Christopher Koper, and Daniel J. Woods. 2011. “A Randomized Controlled Trial of Different Policing Strategies at Hot Spots of Violent Crime.” *Journal of Experimental Criminology* 7: 149–181.

Tyler, Tom. 1990. *Why People Obey the Law.* New Haven: Yale University Press.

Villaveces, Andres, Peter Cummings, Victoria E. Espitia, et al. 2000. “Effect of a Ban on Carrying Firearms on Homicide Rates in 2 Colombian Cities.” *JAMA* 283: 1205–1209.

Weisburd, David, Lawrence Sherman, and Lisa Maher. 1993. “Contrasting Crime-Specific and Crime-General Theory: The Case of Hot Spots of Crime.” In Freda Adler and William S. Laufer, eds., *Advances in Criminological Theory*, Vol. 4. New Brunswick, NJ: Transaction.

Wikstgrom, Per-Olof H. and Vania Ceccato, Beth Hardie and Kyle Treiber. 2010. Activity Fields and the Dynamics of Crime: Advancing Knowledge About the Role of the Environment in Crime Causation. *Journal of Quantitative Criminology* 26 (1): 55–87.

Woods, Daniel J. 2010. PhD dissertation, Department of Criminology, University of Pennsylvania.