IBM Center for
The Business of Government

Improving Performance Series

Five Methods for Measuring Unobserved Events

A Case Study of Federal Law Enforcement

John Whitley
Institute for Defense Analyses

Improving Performance Series 2012

Five Methods for Measuring Unobserved Events: A Case Study of Federal Law Enforcement

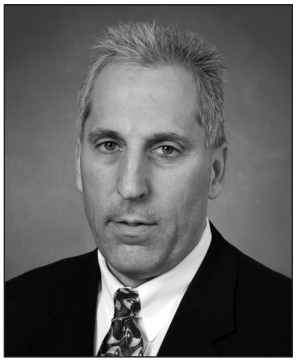
John Whitley
Institute for Defense Analyses

IBM Center for
The Business of Government

Table of Contents

Foreword


Daniel J. Chenok


Gregory J. Greben

On behalf of the IBM Center for The Business of Government, we are pleased to present this report, *Five Methods for Measuring Unobserved Events: A Case Study of Federal Law Enforcement,* by John Whitley.

Measuring program performance is relatively straightforward in many areas of government, such as social services, visa processing, and air traffic control. But in some instances, assessing performance and success is much harder. One particularly difficult area involves law enforcement, where a key goal is to prevent or deter bad outcomes—which can often happen without the knowledge of law enforcement officials.

This report describes the challenges of measuring unobserved events such as tax cheating, drug smuggling, or illegal immigration. These differ from infrequent events (another measurement area where success is difficult to judge), such as terrorist attacks.

The author, John Whitley, is a former executive in the Department of Homeland Security, where he faced these challenges daily. He reviews practical methods already in use in some government and law enforcement agencies to estimate unobserved federal crime rates and ways to measure them. He notes that applying these methods more broadly at the federal level "could enable the types of performance management reforms that have revolutionized local law enforcement."

Whitley describes five data estimation methods: the use of administrative records; surveys; inspections, investigations, and audits; experimental methods; and technical measurement. He provides examples of how these are already being pioneered in areas as diverse as determining the amount of counterfeit money in circulation, the underpayment of taxes, and credit card fraud. He shows how having this kind of information helps law enforcement managers develop appropriate data-driven strategies and approaches for countering these kinds of crimes.

A key insight offered by Dr. Whitley is that there is often "institutional resistance to diverting scarce resources from enforcement to data collection and measurement." But he says that when law enforcement leaders do make the leap, they quickly become

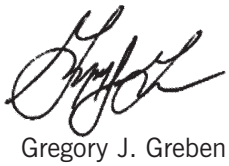
champions of the use of data. Local police departments across the country have begun to use these kinds of techniques to undertake what has become known as "predictive policing"—moving police officers to places where crime is predicted to occur, rather than only responding to crimes that have occurred. This is happening in cities including Los Angeles, Pasadena, and Charleston.

This report is one in a series by the IBM Center exploring more sophisticated approaches to analyzing and using performance information. Other reports on the use of analytics include:

- *Empirically Based Intelligence Management: Using Operations Research to Improve Programmatic Decision-Making* by Frank Strickland and Chris Whitlock
- *From Data to Decisions: Building an Analytics Culture* by the Partnership for Public Service
- *From Data to Decisions: The Power of Analytics* by the Partnership for Public Service
- *Strategic Use of Analytics in Government* by Sirkka Jarvenpaa and Thomas Davenport

We hope this report helps federal law enforcement leaders and other government managers to tackle the tough job of using sophisticated approaches to improve program performance in hard-to-measure subject areas.

Daniel J. Chenok
Executive Director
IBM Center for The Business of Government
chenokd@us.ibm.com



Gregory J. Greben
Vice President, Public Sector
Business Analytics & Optimization
IBM Global Business Services
greg.greben@us.ibm.com

Part I: Introduction

Purpose of this Report

In recent decades the local law enforcement community has been a pioneer in measuring and reporting performance and in using these data to drive strategy development and manage execution. New York City's CompStat[1] revolution focused commanders on "crime trends with the same hawk-like attention private corporations pa[y to] profits and losses. Crime statistics have become the [police] department's bottom line, the best indicator of how police are doing precinct by precinct and citywide" (Smith and Bratton 2001). Although scholars are still debating the relative contribution of performance-driven management reform in the dramatic crime rate decline over the last 20 years, these policing reforms have likely played a significant role and have been emulated around the country in areas well beyond law enforcement.

Law enforcement at the federal level, however, has been slower to adopt these reforms, and many areas of federal law enforcement do not systematically collect, use, or report basic data on crime rates within their jurisdictions. One of the most difficult challenges contributing to this lack of outcome-oriented, data-driven management is how to measure the level of many federal crimes.[2] Federal crimes like drug smuggling and income tax evasion often go undetected—failing to leave a trail of administrative records that produce performance data on crime rates. The perpetrators of these crimes prefer not to get caught, and no directly affected victims exist with the incentive to notify law enforcement officials. Federal law enforcement organizations are left with the challenge of how to measure unobserved events.

These measurement challenges are not unique to federal crimes. No criminal wants to be apprehended. State and local law enforcement organizations struggle with the accurate measurement of crimes like rape, drug use, and prostitution within their jurisdictions—even murder can be surprisingly hard to measure. Similarly, other federal organizations deal with unobserved events in many aspects of their missions as well.

The purpose of this report is to review practical techniques, many already in use in other government and law enforcement areas, that can be used to estimate unobserved federal crime rates and related performance measures. The intended audience includes federal senior executives responsible for managing and allocating scarce resources within law enforcement, along with performance and program management officials tasked with estimating and reporting these measures. Applied to federal law enforcement, the techniques reviewed in this report could enable the types of performance management reforms that have revolutionized local law enforcement. The discussion of these techniques may also be useful in other settings where unobserved events hinder performance management.

1. Crime Statistics, Website of the New York City Police Department. http://www.nyc.gov/html/nypd/html/crime_prevention/crime_statistics.shtml.

2. A crime is an offense against a public law. In its most general sense, it includes all offenses. It can also be used in a more limited sense to only include felony violations of public law. This report uses the word crime in its more general sense to include all types of offenses, whether felonies or not.

Outline of this Report

Part II of this report reviews the basic performance management framework for law enforcement activities. It uses examples to:

- Identify specific, key outcome performance measures essential for sound management
- The challenges in estimating performance
- The problems created when non-outcome measure proxies are used

Part III systematically reviews techniques and methods that can be used to estimate unobserved variables. This includes empirical analysis of administrative records, surveys, and covert testing.

Part IV applies these techniques to specific examples in federal law enforcement to demonstrate how they can be used in practice.

Part V outlines selected major challenges with data quality and interpretation that arise in estimating outcome performance measures based on unobserved events, and includes concluding remarks.

Part II: Performance Management Framework for Law Enforcement

Understanding an organization's performance measurement challenges and developing a plan to overcome them begins with identifying appropriate performance measures. This process, often integral to strategic planning, begins with identifying the outcomes the organization is trying to achieve or influence. For law enforcement organizations, the clear and authoritative place to start is in the laws the organization is tasked with enforcing.

This section defines the primary outcome performance measures for law enforcement and then briefly examines the output, input, and efficiency measures that can be used to effectively implement strategies for the achievement of those outcomes.

Law Enforcement Outcomes

Crime rates, the rates at which laws are broken, are the primary outcomes and outcome performance measures for law enforcement organizations. At the local level, e.g., New York City, the outcome measures for which the chief of police is accountable are the rates of murder, rape, aggravated assault, burglary, etc. City officials above the police chief have broader objectives; for example, in addition to crime, the mayor must focus on the standard of living and economic growth of the city. Officials subordinate to the police chief may have narrower objectives; for example, the manager of the 911 call system may be focused on the response time to emergency calls.

In New York City, the police department measures these outcomes and uses the data to formulate policy objectives, develop and evaluate strategies, monitor execution, and ensure accountability (see O'Connell 2001 and Smith and Bratton 2001 for detailed descriptions of the New York City Police Department's use of performance data and its CompStat system). Figure 1 illustrates a typical publicly available weekly report from the New York City Police Department. This level of transparency with the public has a direct impact on the public debate about crime and expenditures on policing in New York, focusing the debate as much as possible on actual crime rates and how they change as policies and resources are adjusted.

The primary outcomes and, thereby, outcome performance measures for federal law enforcement organizations are the rates at which the federal laws affecting their jurisdictions are broken. For example, the federal government enforces U.S. immigration law, and a primary outcome measure is the rate at which individuals enter and reside in the country unlawfully. Enforcement for most of these laws is the responsibility of the Department of Homeland Security (DHS) and its components:

- Immigration and Customs Enforcement (ICE)
- Customs and Border Protection (CBP)
- United States Coast Guard (USCG)

Figure 1: New York City CompStat Weekly Report

Police Department
City of New York

Michael R. Bloomberg
Mayor

Raymond W. Kelly
Police Commissioner

Volume 18 Number 49

CompStat

112th Precinct

Report Covering the Week 12/5/2011 Through 12/11/2011

Crime Complaints

	Week to Date			28 Day			Year to Date*			2 Year	10 Year	18 Year
	2011	2010	% Chg	2011	2010	% Chg	2011	2010	% Chg	% Chg	% Chg (2001)	% Chg (1993)
Murder	0	0	***.*	0	0	***.*	2	1	100.0	0.0	***.*	-60.0
Rape	0	0	***.*	1	1	0.0	11	4	175.0	83.3	266.7	10.0
Robbery	2	2	0.0	9	7	28.6	100	113	-11.5	-7.4	-49.7	-83.7
Fel. Assault	0	2	-100.0	3	5	-40.0	58	56	3.6	-3.3	-38.3	-52.5
Burglary	5	5	0.0	13	13	0.0	135	125	8.0	-19.6	-59.8	-89.7
Gr. Larceny	12	9	33.3	38	27	40.7	457	365	25.2	18.7	-27.5	-52.4
G.L.A.	3	1	200.0	11	8	37.5	101	96	5.2	13.5	-82.9	-96.8
TOTAL	***22***	***19***	***15.79***	***75***	***61***	***22.95***	***864***	***760***	***13.68***	***5.62***	***-53.35***	***-86.13***
Petit Larceny	21	20	5.0	72	72	0.0	1,049	929	12.9	60.4	***.*	***.*
Misd. Assault	6	4	50.0	15	19	-21.1	178	209	-14.8	-2.2	***.*	***.*
Misd. Sex Crimes	0	0	***.*	0	0	***.*	16	15	6.7	-27.3	***.*	***.*

Historical Perspective

(Historical perspective is a complete calendar year of data.)

	1990	1995	1998	2001	2010	%Chg '10 vs '01	%Chg '10 vs '98	%Chg '10 vs '95	%Chg '10 vs '90	
Murder	7	2	2	0	1	/0	-50.0	-50.0	-85.7	**Murder**
Rape	9	7	7	3	4	33.3	-42.9	-42.9	-55.6	**Rape**
Robbery	698	453	272	213	119	-44.1	-56.2	-73.7	-83.0	**Robbery**
Fel. Assault	137	135	84	98	58	-40.8	-31.0	-57.0	-57.7	**Fel. Assault**
Burglary	1,680	769	605	350	137	-60.9	-77.4	-82.2	-91.8	**Burglary**
Gr. Larceny	1,340	789	741	668	389	-41.8	-47.5	-50.7	-71.0	**Gr. Larceny**
G.L.A.	3,477	1,973	925	622	99	-84.1	-89.3	-95.0	-97.2	**G.L.A.**
TOTAL	***7,348***	***4,128***	***2,636***	***1,954***	***807***	***-58.70***	***-69.39***	***-80.45***	***-89.02***	***TOTAL***

The above CompStat figures are posted on Monday, one week after the closing date.

CompStat figures are preliminary and subject to further analysis and revision. Crime statistics reflect New York State Penal Law definitions and differ from the crime categories to the F.B.I. Uniform Crime Reporting System. All degrees of rape are included in the rape category.

Prepared by
NYPD CompStat Unit

CompStat

The senior officials in these organizations are responsible for enforcing U.S. immigration laws. Among the outcomes for which they should be held accountable are the rates at which U.S. immigration laws are broken. The President, and members of White House policy councils, may also have broader objectives, such as economic growth and security, which are affected by violations of immigration law. Subordinate program offices within CBP, ICE, and USCG may also have narrower objectives, such as situational awareness within a particular border domain and the time it takes to process a detained illegal immigrant for deportation.

Unlike robbery and assault, which leave aggrieved victims with an incentive to notify the police and can involve abundant physical evidence, violations of immigration laws are largely unobservable to federal law enforcement officials unless the perpetrators are apprehended in the act. Estimating the rate of illegal immigration is essential for data-driven management in

immigration law enforcement. Improving the availability of objective and transparent performance outcome data would improve management and decision-making in areas like:

- **Setting clearer policy objectives:** Many areas of federal law enforcement, like border control and drug trafficking regulation, are politically contentious. Providing objective data on outcomes can help focus debate on choices between fact-based alternatives, encouraging the establishment of actual objectives or, at least, the concrete identification of where agreement has been reached on objectives and where it has not.
- **Selecting the right strategies:** Because of the political nature of the issues and the failure to arrive at clear objectives, choices among strategies often become contentious in federal law enforcement. For example, policy-makers may choose the number of border patrol agents and the types of employment immigration enforcement based on a desire to demonstrate strength, rather than rigorous empirical evidence of the strategy's effectiveness when compared to alternative enforcement tools. Objective measurement of outcomes and the performance levels of alternative strategies help to focus the selection of strategies on their actual contributions to outcomes instead of on the political signals they send.
- **Managing execution:** With clear objectives and an assessment of how strategies are working, adjustments in execution can be defended politically by providing evidence to support those adjustments.
- **Ensuring accountability:** When agencies and agents are given clear objectives and there are measurement strategies in place to determine if the objectives are achieved, execution can be decentralized and accountability maintained by measuring results produced instead of inputs expended, e.g., budget spent or number of agents deployed. Congress and the Administration may be more willing to focus on setting objectives and evaluating strategies, rather than micromanaging execution, because they know that management officials can be held accountable for the resources they have been provided.

When outcomes are hard to measure, proxy variables are often used, but they can be ineffective or, worse, counterproductive. A common proxy measure in law enforcement is the number of arrests for various crimes. But as Washington, D.C. Police Chief Cathy Lanier said recently, "Arresting people is not a measure of success. Less crime is a measure of success" (Goldberg 2012). Arrests are an important element in the production of law enforcement that should be measured, but they are not the outcome, and they cause confusion when they are reported and used as such. In immigration law enforcement, one equivalent proxy measure is the number of apprehensions of illegal immigrants attempting to cross the border. But, as reported in a recent study, "CBP [attributed] *increases* in apprehensions made at checkpoints in some border sectors to improved operations and *decreases* in apprehensions in other sectors to the deterrent effects of improved CBP technologies and increased staffing. Clearly, a measure that reflects successful performance whether it rises or falls has limited value as a management tool" (RAND 2011).

In addition to reducing accountability and failing to provide management with useful tools, measures that do not act as a true proxy for outcomes can focus attention and resources on the wrong things. This becomes counterproductive to achieving the desired outcomes. Former Mayor Rudolph Giuliani has stated that, in New York City, he found that focusing on arrest rates actually distracted police officers from a focus on crime:

> We were equating success with how many arrests were made. A police officer was regarded as a productive police officer if he made a lot of arrests. He would get promoted. A police commander in a precinct would be regarded as a really good police commander if his arrests were up this year. This wasn't the only measure of success, but it was the predominant one.

> Arrests, however, are not the ultimate goal of police departments or what the public really wants from a police department. What the public wants from a police department is less crime. So it seemed to me that if we put our focus on crime reduction and measured it as clearly as we possibly could, everybody would start thinking about how we could reduce crime. And as a result, we started getting better solutions from precinct commanders (Giuliani 2000).

Producing, reporting, and explaining proxy measures also consume resources, making those resources unavailable for use in measuring actual performance outcomes. In addition to reporting apprehensions, DHS reported the measure Border Miles Under Effective Control for several years. In 2006, a multi-billion-dollar, five-year investment plan was launched for the southwest land border, and DHS intended to use the measure to evaluate the effects of the plan; however, the measure was never objectively defined. By not capturing meaningful outcomes, this measure introduced confusion and unnecessary contention into the public debate instead of grounding debate in facts and evidence.

The measure was ultimately abandoned, leaving DHS with limited ability to evaluate the effectiveness of the border buildup or make subsequent decisions about how to execute the initiative. The likely costs of the types of measurement proposed in this report would not exceed, and could even be less than, the amount DHS spent on the Effective Control measure. DHS has stated that it is now creating a new set of proxy measures that will be reported as an index. The resources being committed to this effort might be better spent trying to measure actual outcomes of interest.

Supporting Performance Measures

Although the focus of this report is primarily on measuring unobserved law enforcement outcomes, a complete performance management framework requires measurement of more than just outcomes. Law enforcement activities affect crime rates in a variety of ways. To effectively manage activities so as to influence outcomes, law enforcement officials need to understand and measure how their activities affect outcomes. In addition to outcome performance measures, a law enforcement performance management framework needs to include:

- Measurement of outputs, inputs, and efficiency
- Empirical analysis on how these measures relate to each other and to outcomes

Law enforcement activities reduce crime in two primary ways:

- Stopping crimes (during planning or in the act)
- Deterring crimes from occurring in the first place

In much of law enforcement, the primary method of reducing crime is through deterrence, as few crimes are actually stopped in the act. Federal law enforcement, however, offers some examples where stopping a crime in the act is a primary role of law enforcement, e.g., terrorism and drug smuggling. Deterrence is often further broken down into two elements:

- **Specific deterrence** focuses on an individual who is committing a crime; arresting and punishing the individual may prevent a repeat of the offense
- **General or indirect deterrence** refers to the general prevention of crime by deterring potential criminals from committing a crime in the first place

For example, in border law enforcement, these elements have been called *at-the-border deterrence*—preventing an illegal immigrant from repeating an attempted border crossing—and

behind-the-border deterrence—stopping a potential illegal immigrant from making the initial decision to attempt migration and traveling to the border. This deterrence is produced most directly by arresting and punishing offenders, and letting other potential offenders know it will happen to them.

This understanding helps to identify the first level of supporting measures and analyses needed for data-driven management in law enforcement:

- The number of arrests and the subsequent arrest rate (arrests per crime committed)
- The rates at which these individuals are charged and convicted[3]
- The punishments administered

A key analysis at this level is the relationship between these variables (collectively identified as the expected consequence of committing the crime) and the level of crime, i.e., the level of deterrence created. In areas of law enforcement with a high volume of observed crime that responds quickly to changing circumstances, this analysis might be done in real time through varying enforcement efforts. In other situations, such as illegal immigration, it may require detailed empirical studies conducted over long periods of time. This analysis can be challenging and may involve disentangling interrelated effects—e.g., the effect of law enforcement and economic changes on illegal immigration rates—but is necessary for effective management.

At this level of supporting measures, there may be other important issues that also need to be considered. For example, in law enforcement related to drug use, rehabilitation may be considered a substitutable (or complementary) consequence and, in this case, would also have to be measured. Analysis would then be used to determine its effectiveness at preventing crime as well as its effectiveness relative to punishment.

The next level of supporting measures and analyses identifies how the key outputs are produced. Arrests are produced by intelligence gathering, patrolling, investigating, and numerous other activities. Identifying the available tools and their relative contributions to making arrests is necessary for optimally allocating resources to efficiently increase arrest rates. Charges and convictions are achieved through further investigation, evidence gathering, and legal analysis. Punishment consists of fines, incarceration, and other legal sanctions.

An effective performance management framework should be built around a detailed understanding of how law enforcement activities contribute to the outcomes the officials are trying

The Need for More Academic Research

Extensive academic literature in economics and criminology empirically measures the effects of arrests, prosecutions, and punishments on crime rates. This literature helps law enforcement officials and policy-makers prioritize scarce enforcement resources, but much of the available literature is focused on crimes of state and local primary jurisdiction. The lack of analysis on crimes of primarily federal jurisdiction is driven in part by the lack of systematically reported data. Improved performance reporting in federal law enforcement could help create an academic literature that would assist federal decision-makers in effectively prioritizing investments and managing law enforcement functions.

3. For some areas of federal law enforcement, e.g., border control, the apprehension rate is equivalent to the rate for both arresting and charging. In general law enforcement, this is sometimes called the clearance rate.

to achieve. This requires both measuring these activities and analyzing their relationships to each other and to outcomes. Some of this analysis may be possible to do informally during execution of actual operations, and some may require more detailed empirical analysis, but effective stewardship of taxpayer resources requires that it be done.

Part III: Five Methods for Estimating Data on Unobserved Events

Tax evasion or illegal drug smuggling are often not observable events for federal law enforcement officials. But to effectively manage federal law enforcement activities, officials and policy-makers in charge must have an idea of what is happening. Five methods that can assist government performance analysts in estimating basic information on unobserved events are introduced and described here.

The Need for a Statistical Framework

Law enforcement can face tough measurement challenges, but the fields of statistics and econometrics have developed a framework for dealing with them and it is useful to begin this part with a brief overview of that framework. All violations of a federal law can be thought of as elements of a prospective data population. The scope of the population can be defined in various ways—e.g., immigrants illegally entering the United States in a calendar year, or the illegal drugs smuggled across the southwest land border between the United States and Mexico. To effectively manage their operations, federal law enforcement officials need insight into these unobserved violations; i.e., they need to know the properties or parameters of this population of data, such as its size and distribution.

Law enforcement officials are generally able to observe subsets, or samples, of this population. The most obvious is the subset of violators apprehended or arrested. Detailed documentation of apprehensions or arrests is generally retained in administrative records. In addition, there may be other available sources of data, often partial and incomplete, that shed light on various aspects of the population, e.g., survey data on drug usage or the footprints in the desert of illegal border-crossers. Actions can also be taken to increase the available data, such as increasing the size of the observable subset, drawing additional samples from the population, or generating a sample of new data that mimics the characteristics of the population of interest. The methods described here use such samples to make estimates of the total population.

When using a sample to estimate parameters of the underlying (unobserved) population, an important statistical property is whether the estimate is biased. Bias occurs when the estimate systematically diverges from the true value of the population parameter being estimated. An unbiased and therefore preferred estimate does not systematically diverge from the true value. One primary cause of bias is a poor sample that is not randomly selected. A sample is random when every element of the population has an equal probability of being included. Examples of non-random samples may include:

- Records on individuals apprehended smuggling drugs across the border. These records may not be representative of all individuals who attempt to smuggle drugs across the border if slower, less prepared individuals are more likely to be caught.

- Survey data on the propensity to illegally migrate to the United States collected from urban Mexican households. These data may not be representative of the propensity of all Mexicans, both rural and urban, to illegally migrate.

In these cases, estimates of population parameters made from the sample data may be biased and thus misleading. It is important for government managers who develop performance measures to be constantly vigilant for bias in their estimates. It may not be possible to eliminate all potential biases in data, but the analyst must be aware of the major potential biases in their data and their possible effects.

A final note on the need for a statistical framework in the area of law enforcement: although often related, the challenges of measuring *unobserved* events are different from the challenges associated with *infrequent* events. The Department of Defense prepares to fight wars, but fortunately these are very infrequent. DHS prepares for a nuclear attack on a major U.S. city, but fortunately this has never happened. Measuring the performance of military capability in a war fight when there is no war fight or the performance of response and recovery capabilities for a terrorist nuclear attack when there has never been one are very important, but are not the focus of this report.

Five Data Estimating Methods

Method One: Administrative records. Once a performance manager has identified the outcomes that need to be measured and is beginning the task of developing a measurement strategy, the first action is to identify all relevant data currently captured by the agency or by others. In the best-case scenario, the performance manager may discover relevant data at a lower level in the organization (e.g., at the field offices) or in another organization (e.g., in a survey conducted by the Census Bureau that asks a pertinent question).

Or it could be that estimation of the outcome is possible, but that multiple sources of data have to be combined and those sources are spread across organizations. For example, Immigration and Customs Enforcement is responsible for law enforcement concerning individuals who enter the United States on visas, but violate the visa by overstaying the required departure date. The rate of visa overstay, however, is unobservable to federal law enforcement. The number of visas issued and their required departure dates are known, but who actually departs and when is not. Everyone leaving the United States by commercial air or maritime transport is known because they are identified in passenger manifest documents maintained by the transportation companies.[4] Thus, a major portion of the performance measure can be estimated by combining data from government visa records with commercial transportation passenger manifests.

Inventorying available data is a valuable exercise even when it does not identify a previously unrealized, accessible method of estimation, because it may identify ways to facilitate estimation by enhancing existing data collection. Examples might include the addition of another information field into a case management IT system or the addition of a question to an existing government survey. In these cases, estimation is possible but the required data are simply not being captured.

4. These outbound passenger manifests are an incomplete data source, however, because they are not available for land border exits and may not be complete for non-commercial air and maritime exit. Therefore, although a major set of data that can be used in computing a visa overstay rate is available from existing data spread across multiple sources (some of which are outside of the government), this illustrates another important point—it may be necessary to combine methods to develop complete measures. In this case, the use of outbound passenger manifests may have to be combined with techniques identified below to develop a complete visa overstay performance measure.

Example of Using Administrative Records Recidivism Analysis

The U.S. Border Patrol (USBP), within the DHS Customs and Border Protection, is responsible for controlling the U.S. land border with Mexico between the Ports of Entry (POEs). USBP maintains a detailed database called ENFORCE on all apprehensions of illegal border-crossers. When illegal border-crossers from Mexico are apprehended and returned to Mexico, many try again within a relatively short period of time. In fact, if all who were returned attempted to cross again, the fraction apprehended a subsequent time would constitute an estimate of the apprehension rate. With an estimated apprehension rate in hand, it is then possible to estimate the flow of illegal border-crossers.

A challenge with this approach is that every individual who is apprehended and returned does not attempt a subsequent crossing—the first apprehension acts as a deterrence (referred to as at-the-border deterrence in the earlier discussion on supporting measures). This means that recidivism analysis by itself does not solve measurement challenges, but it can provide an important part of the solution in situations where it is appropriate. If it can be combined with estimates from other sources on the deterrence effect of apprehensions, it can be used to create an estimate of apprehension rate and, subsequently, the rate of illegal immigration.

In the absence of data from other sources on the deterrence effect of an apprehension, an additional technique that can be used to increase the usefulness of recidivism analysis is sensitivity analysis. The performance manager could estimate the implied apprehension rate from the recidivism data, assuming different levels of deterrence. The analysis could start with zero percent deterrence (the assumption in the motivating example above) and then consider 10 percent, 20 percent, etc., where 10-percent deterrence means that 10 percent of individuals apprehended at the border are deterred from making a subsequent attempt. This analysis would provide a range of possible estimates depending on the level of deterrence.

Another potentially useful result from inventorying the available data is to identify data—or new data that could be generated—that could, with the use of clever empirical methods, support analyses that would estimate the outcomes of interest. In the *Recidivism Analysis* box, a method for using apprehension data to infer an apprehension rate, and subsequently the rate of illegal immigration, is presented.

Method Two: Surveys. Surveys are a commonly used data collection method in policy and social science research. Surveys involve asking a set of questions to a sample population. They can be conducted by telephone or mail, online, or in person. The goal is to obtain a sample of sufficient quality, e.g., size and representation, to enable inferences to be drawn about the population from analysis of the data. Surveys may be conducted on a regular, recurring basis to create estimates through time or can be conducted on a one-time basis to answer specific questions at a point in time.

There are numerous surveys already being conducted by the government and private organizations that provide valuable information on federal law enforcement issues. The U.S. Bureau of the Census and its many supporting surveys provide some of the most comprehensive data about the United States. Other federal agencies conduct a wide range of surveys that include specific emphasis on law enforcement issues, such as the National Crime Victimization Survey (NCVS) discussed in the *National Crime Victimization Survey* box that follows. Surveys are

Example of Using Surveys
National Crime Victimization Survey (NCVS)

Many state and local crimes leave aggrieved victims and physical evidence and are more likely to be reported to law enforcement officials than some federal crimes are. But this reporting is still not perfect, and the records of these reported crimes may not represent the full extent of the crime committed. The NCVS is one method law enforcement officials use to understand and measure potential undercounting.

According to its official website, the NCVS surveys a nationally representative sample of about 40,000 households on criminal victimization in the United States. Each household is interviewed twice during the year. The data are then used to estimate the likelihood of victimization by rape, sexual assault, robbery, assault, theft, household burglary, and motor vehicle theft for the population as a whole as well as for segments of the population such as women, the elderly, members of various racial groups, city dwellers, or other groups.

More information is on the NCVS is available from the Department of Justice at http://bjs.ojp.usdoj.gov/index.cfm?ty=dcdetail&iid=245.

also conducted by academic researchers, think tanks, and private companies. In some situations, there may already be a recurring survey conducted that is close to, but not exactly, what the performance manager needs; a cost-effective way to get started is to partner with the organization conducting the existing survey to expand it in a way that would be useful for the law enforcement performance measurement.

There are, of course, challenges and limitations to surveys. It can be costly to implement a new survey that contacts enough individuals to be statistically valid. It can also be challenging to elicit truthful answers, particularly on issues of interest to law enforcement like criminal activity. There are generally few consequences for responding untruthfully on a survey, and on questions related to criminal activity, fear of how the information will be used may be a strong incentive to lie. A great deal of work has been done to structure questions and surveys to elicit truthful answers and test for inconsistencies within a survey response, and the results allow this risk to be mitigated, but it remains an issue when considering the use of surveys.

When the decision has been made to further explore use of a survey to address a measurement challenge, important factors that will have to be addressed include the survey methods (e.g., telephone, one-on-one), the sampling methodology (e.g., random drawing of names, targeting of particular groups), question design (must be unambiguous, must allow estimation of the desired measure), question sequencing, starting with a pilot or the full survey, and how the analysis will be conducted once the survey is complete. Numerous professional firms that conduct surveys and academic sources of information can be explored further.

Method Three: Inspections, Investigations, and Audits. Criminal or administrative investigations offer another way to systematically collect an accurate data sample. The important point about using investigations in the context of measuring unobserved events is that the investigations must be in some way random. In typical law enforcement operations, proactive investigations are prioritized to follow the most important clues or those that are most likely to lead to a major arrest or disruption of crime. Investigations prioritized in this manner may not provide statistically valid estimates of the underlying level of criminal activity. Conducting investigations

Example of Using Audits
National Research Program

The Internal Revenue Service (IRS) within the Department of the Treasury is the nation's tax collection agency and administers the Internal Revenue Code. The *tax gap* is the IRS's measure of tax liability that is not paid on time. The IRS National Research Program (NRP) measures the tax gap using randomized audits.

Most IRS tax audits are targeted to those tax returns for which there is suspicion of non-compliance. These audits cannot be used to develop an estimate of overall compliance because they are not a representative sample of all tax returns. The NRP, therefore, conducts audits on a random set of tax returns to develop an unbiased estimate.

The NRP originally drew samples every few years of about 45,000 individuals. In 2007, it switched to an annual sampling of 13,000 individuals. This allows the IRS to make more frequent estimates and more accurately monitor trends. The majority of individuals selected will have their tax returns confirmed through in-person audits with an IRS examiner. The IRS will also use matching and third-party data to confirm the accuracy of the tax returns.

In addition to measuring the tax gap, the IRS also uses the NRP to update and improve the accuracy of its audit selection tools. By identifying the correlations between reported data and non-compliance in the randomized sample of tax returns, the IRS is able to refine the set of factors it looks for in all tax returns to prioritize auditing. In other words, collecting data and measuring performance not only help in strategic analysis such as evaluating the level of investment the United States should make in tax enforcement—they also help optimize the range of tools the IRS employs to ensure tax law compliance.

Another important thing to note about the NRP is that randomized audits only measure compliance for individuals who actually file a tax return. Another form of noncompliance is to not file a return at all. The NRP uses surveys and other tools to measure this form of noncompliance. This provides another example of multiple methods being combined to develop a complete estimate.

on a more random sample of potential illegal activity represents a major cultural shift for law enforcement operations, but limited and systematic use of them can be a powerful way to collect information about the outcomes the law enforcement organization is trying to effect. See the *National Research Program* and the *Administrative Site Visit and Verification Program* text boxes for examples of this method.

Method Four: Experimental Methods. Another method involves actually adding or modifying law enforcement activities in the field in ways that may facilitate estimation of the crime rate. In controlled environments like Ports of Entry or airport security screening, this could involve selecting a randomized subset of individuals who pass the primary screen for a secondary, more rigorous screen. The rate at which violations are identified in the secondary screen can be used to infer the failure rate of the primary screen. The *Randomized Secondary Screening* text box describes how CBP conducts these randomized secondary inspections at Ports of Entry. This method is not restricted to physical screening—application processing and other forms of information-based screening can also have randomized secondary evaluations conducted to evaluate the accuracy of the primary screening process.

Example of Using Inspections
Administrative Site Visit and Verification Program (ASVVP)

The U.S. Citizenship and Immigration Service (USCIS) within DHS is the government organization that oversees lawful immigration to the United States. Detecting and preventing the fraudulent obtainment of an immigration or citizenship benefit is an important area of federal law enforcement. An example of such fraud would be the establishment of a phony religious organization created to sponsor individuals for immigration using a religious worker visa. This was a particular concern following the terrorist attacks of September 11, 2001.

The USCIS program ASVVP uses systematic site inspections to verify information contained in visa petitions. Site inspections are unannounced and can be conducted pre- and post-adjudication. When potential fraud is identified, the evidence is turned over to law enforcement officials for potential criminal investigation.

Significantly, the site inspections are not performed only where fraud is suspected, but are randomly selected so that a sample of all petitions is created and an unbiased estimate of fraud can be performed. USCIS uses this program to estimate performance measures for fraud and reports them in the DHS Annual Performance Report.

Example of Using Field Experiments
Randomized Secondary Screening

The Office of Field Operations within U.S. Customs and Border Protection (CBP) is responsible for screening all individuals entering the United States at Ports of Entry. With 340 million individuals entering the United States at these ports per year, this is a high-volume process. All individuals are subjected to a primary screening procedure to ensure compliance with U.S. entry law.

To empirically estimate the failure rate of the primary screening process, CBP randomly selects a sample of the entrants at both air and land ports to conduct a more thorough examination for major violations. Major violations involve serious criminal activity, such as possession of narcotics, smuggling of prohibited products, human smuggling, weapons possession, fraudulent U.S. documents, and other offenses serious enough to result in arrest. For the air domain, passengers are selected in a random sample that totals 12,000 passengers annually (1,000 passengers per month) at each of the 19 largest international airports. Similarly, for the land domain, passengers are selected in a random sample that totals 12,000 passengers annually (1,000 passengers per month) at each of the 25 largest land border ports. These sample sizes were selected to obtain an overall 95-percent confidence level in the estimates.

Another example might be the randomized surge of enforcement activity across different geographic regions. The U.S. Border Patrol could conduct an unannounced surge of agents into a randomly selected station and measure the increase in drug apprehensions. Conducting these random surges across a range of locations and circumstances could be used to infer information about the population of offenders who are evading detection and apprehension during non-surge periods. In addition to performance measurement, these random surges can also provide a deterrent effect to crime and have been used throughout law enforcement. For example, the Amtrak Police Department uses random and unpredictable surges as part of its law enforcement strategy.

Another category of experimental methods is to conduct controlled tests. With this method, the organization is not collecting data about the underlying population; it is generating a new data population in a controlled way so that this new population has the same properties as the population of interest. Perhaps the most prominent method in this category is red-teaming, also called covert testing or penetration testing. A red team is an independent group that seeks to challenge an organization for the purpose of detecting vulnerabilities. The *Red-Teaming* text box describes the use of this method in aviation security, where penetration testers physically test security. Its most obvious application to law enforcement performance measurement is in situations where there is a controlled environment such as inbound passenger screening at POEs. In less controlled environments, e.g., drug smuggling routes along the border, there are obviously safety issues for the testers that would have to be seriously evaluated before considering this method.

Red-teaming does not have to be physical penetration testing, however. Many federal law enforcement organizations and their counterpart benefit-delivery organizations (e.g., the

Example of Using Experimental Methods Red-Teaming

Red-teaming, also known as covert testing, is done throughout the federal government; examples range from the Government Accountability Office to the National Nuclear Security Administration. The Transportation Security Administration (TSA), within DHS, conducts systematic red-teaming of passenger and baggage screening in aviation travel. Some of this red-teaming is done to explore new techniques that terrorists might use to identify gaps and vulnerabilities in current systems. More important for purposes of this report, however, is the other portion of TSA red-teaming that is done explicitly for the measurement of screening performance—the failure rate in detecting threats such as firearms, knives, Improvised Explosive Devices, and emerging threats.

This performance testing is part of TSA's Aviation Screening Assessment Program (ASAP). Although the actual operational details of this program are classified, TSA has released to the public some basic information. Each airport receives a prescribed number of assessments that they are required to conduct within a six-month cycle. The tests are unannounced and conducted surreptitiously; i.e., they are covert tests. The number of tests conducted is large enough to ensure that the sample collected can yield statistically significant estimates of the failure rate. Test result data are standardized and maintained in a centralized database to facilitate performance measure estimation and other empirical analysis. Through this program, TSA is able to compute statistically valid estimates of its failure rates in both passenger and baggage screening.

USCIS) engage in information-based screening as well as physical screening. In an information-based screening context, where data are validated to ensure compliance with the law (e.g., visa and citizenship application processing), red-teaming could entail the systematic filing of phony applications to identify failure rates of the screening process. These methods are similar to mystery shopper testing done by the private sector to test service in retail outlets.

Red-teaming aims to conduct controlled tests, but is performed directly in the field (covertly) with the operational activities as they execute their daily mission. An even more controlled test is one that takes place in an artificial environment; e.g., a laboratory or range test. This is done routinely across the government in acquisition programs where both development and operational test and evaluation are standard procedures in the procurement of newly designed complicated government hardware such as defense weapon systems. It can also be used as part of a systematic performance measurement process.

For example, the U.S. Border Patrol uses sensors and radars along the border to detect illegal border crossing. Extensive testing is performed on these items during procurement to understand their false negative and positive rates; that is, the rate at which they miss finding an item of interest and the rate at which they report finding something that is not of interest. Although actual field conditions will differ from these controlled conditions, it may be possible to use the results of these controlled tests to infer likely characteristics of the system in the field.

Method Five: Technical Measurement. Although there are many more methods that can be used, the final method described here is technical data collection. Well-known examples at

Example of Using Technical Measurement Counterfeit Detection

The original mission of the U.S. Secret Service, now within DHS, was to investigate counterfeiting of U.S. currency. Although presidential protection was later added and is now what the Secret Service may be best known for, the Secret Service remains the primary law enforcement organization on counterfeiting. Working with the Federal Reserve Board (FRB), the Secret Service is able to estimate the level of counterfeit currency in circulation in part through technical measurement.

Most currency in circulation gets deposited with financial institutions relatively frequently throughout its circulation life. For safekeeping, space, and financial reasons, these financial institutions send much of this currency to the Federal Reserve Bank for the district in which the financial institution resides. The Federal Reserve Banks then use high-speed currency processing machines to verify the deposits, for both authenticity (counterfeit) and fitness (worn or damaged). For example, the Federal Reserve Bank of New York processes more than 19 million notes each business day. Unfit notes are destroyed (the Federal Reserve Bank of New York destroys approximately five million notes per business day) and counterfeit notes are turned over to the Secret Service.

From this, the FRB can compute a counterfeit rate per one million notes *processed.* This is not a complete measure, however, because counterfeiting, particularly lower quality counterfeiting, is often detected before the currency makes it to the Federal Reserve Banks. To control for this, the Secret Service combines the counterfeit notes detected by the FRB with notes turned in from other sources and law enforcement activity. The Secret Service then uses this total amount of counterfeit passed to compute a measure of the percentage of counterfeit as compared to genuine U.S. currency in circulation.

Example of Using Technical Measurement
Measuring Drug Production

A major area of federal law enforcement is combating the smuggling of illegal drugs into and within the United States. Although some drugs have significant domestic production, e.g., methamphetamine, many drugs are predominantly produced internationally and smuggled into the United States, e.g., cocaine. Identifying the flow of illegal drugs into the United States is an important measure, but unobserved to federal law enforcement officials.

The U.S. Government and the United Nations both produce systematic estimates of drug flows and these estimates start with technical measurement by satellite and aerial imagery. For cocaine, the U.S. estimates are produced by the Inter-Agency Assessment of Cocaine Movement (IACM). These estimates use Intelligence Community (IC) imagery of coca-producing countries to estimate the total level of cultivation. Subsequent analyses include likely harvest yields, refined product yields, distribution destinations (i.e., how much goes to U.S. markets versus markets in other countries), and flow across individual vectors or pathways (e.g., overland through Mexico versus maritime transit through the Caribbean). The final estimates are thus produced by combining many of the different methods described in this report, but the estimation starts with technical collection by satellite and other imagery.

the state and local level include red-light and speeding cameras and, more recently, gunfire detectors in some major cities. Examples at the federal level include the use of sensors, radars, and unmanned aerial vehicles to detect illegal immigrants crossing the border, and radiation detectors and X-ray screening of containerized cargo entering the United States. See the *Counterfeit Detection* and the *Measuring Drug Production* text boxes for examples of this method.

This method has a wide range of potential costs. While a red-light camera may be relatively inexpensive (and quickly pay for itself in fine revenue), arraying sensors and radars across the 1,900 miles of southwest land border is very expensive. When considering this method, the performance manager should consider if there are ways that the unobserved crimes can be observed by technical means and, if so, whether the cost justifies the benefit.

Selecting Between and Implementing the Methods

There is no simple formula that identifies the right methods for performance managers to use. Each measurement challenge is different, and the best methods vary with the circumstances of the individual challenge. There are some rules of thumb, however, that can add structure to the selection process.

- First, the performance manager should identify all relevant data currently captured by the agency and elsewhere. This may lead to a discovery of methods for direct estimation from available data or by expanding existing data sources.
- A second step is to thoroughly examine the context and setting within which the law enforcement activities and performance measure challenges occur. Is there an entity, whether governmental or not, with an incentive to report or measure the crime? If there is an entity suffering an economic loss from the crime (such as credit card companies

and banks discussed in the *Credit Card and Bank Fraud* text box), they may already be measuring it. There may also be think tanks, interest groups, or academic organizations that for various reasons have decided to measure the level of crime. Can measurements from these organizations be validated and used?

When data that allow for direct measurement are not discovered, and no other organization can be found that is already measuring or can be persuaded to measure the offense rate, it is time to begin exploring the estimation methods identified above, and any additional methods that may be relevant. All potentially relevant methods should be identified and then systematically evaluated for their costs and benefits, e.g., how precise the estimates will be, how disruptive to current operations collection will be, and whether they can be implemented in a repeatable and transparent manner. The relevant methods can then be compared to select the best candidates for further development. This shorter list of the best candidates should then be developed more thoroughly to decide which ones should be fully implemented and how. It may also take some trial and error or pilot projects to finalize selection. It also may take the combination of two or more methods to develop a complete estimate for a particular performance measure.

Technical expertise may need to be obtained. The performance management office may want to hire a statistician, economist, or similar technical expert to conduct the work in-house or to oversee supporting work being done by outside entities.

Example of Implementing Methods
Credit Card and Bank Fraud

The United States Secret Service is responsible for investigating violations of federal law with respect to credit card and bank fraud, e.g., obtaining and using a credit card under a false identity without paying the bill. The Secret Service does not directly observe the level of credit card and bank fraud, and developing its own methods to estimate it would be challenging and costly. But victims of these crimes—the credit card companies and banks against which the fraud is conducted—suffer significant losses from their occurrence. Not surprisingly, therefore, credit card companies and banks do try to estimate the levels of fraud, and they have direct access to the data required for measurement because they own it.

Each company has its own data and estimates that could be systematically collected and combined to create aggregate estimates. In many cases, the companies have already done this through trade organizations. For example, organizations such as the Association for Financial Professionals and the American Bankers Association survey companies to measure loss due to fraud (credit card and check fraud), the frequency and nature of fraud, and the environments in which fraud occurs on an annual or periodic basis. Using this information, these organizations make suggestions regarding best practices in fraud protection and identify the leading threats to banks and businesses in the fraud market. Other organizations like Javelin Strategy & Research and the Consumer Sentinel Network (a part of the Federal Trade Commission) also survey consumers and law enforcement agencies regarding fraud loss and reporting.

Utilizing these preexisting resources to estimate the amount of credit card and bank fraud may be substantially cheaper than creating a new system to track and measure the amount of fraud per year nationwide.

Part IV: Applying the Methods: A Case Study of Measuring Illegal Immigration

The methods described in Section III, and other methods not discussed, can be used to estimate unobserved crime rates in a wide range of federal law enforcement settings. To illustrate how the methods might be combined in an actual agency, this section develops a comprehensive example using one of the largest law enforcement areas of the federal government—enforcement of U.S. immigration laws.

Enforcement for most U.S. immigration law is the responsibility of DHS and its components ICE, CBP, and the USCG. The primary violations are being in the United States unlawfully and, for the border enforcement components of DHS (CBP and USCG), the act of entering unlawfully. Entry of illegal immigrants can be subdivided into whether it occurs at a U.S. Port of Entry (POE)—for example, attempting to enter with falsified documents or in the trunk of a car—or between POEs. CBP is responsible for control of entry at the POEs, while responsibility for entry between the POEs is divided between CBP (generally the air and land domains) and USCG (generally the maritime domain).

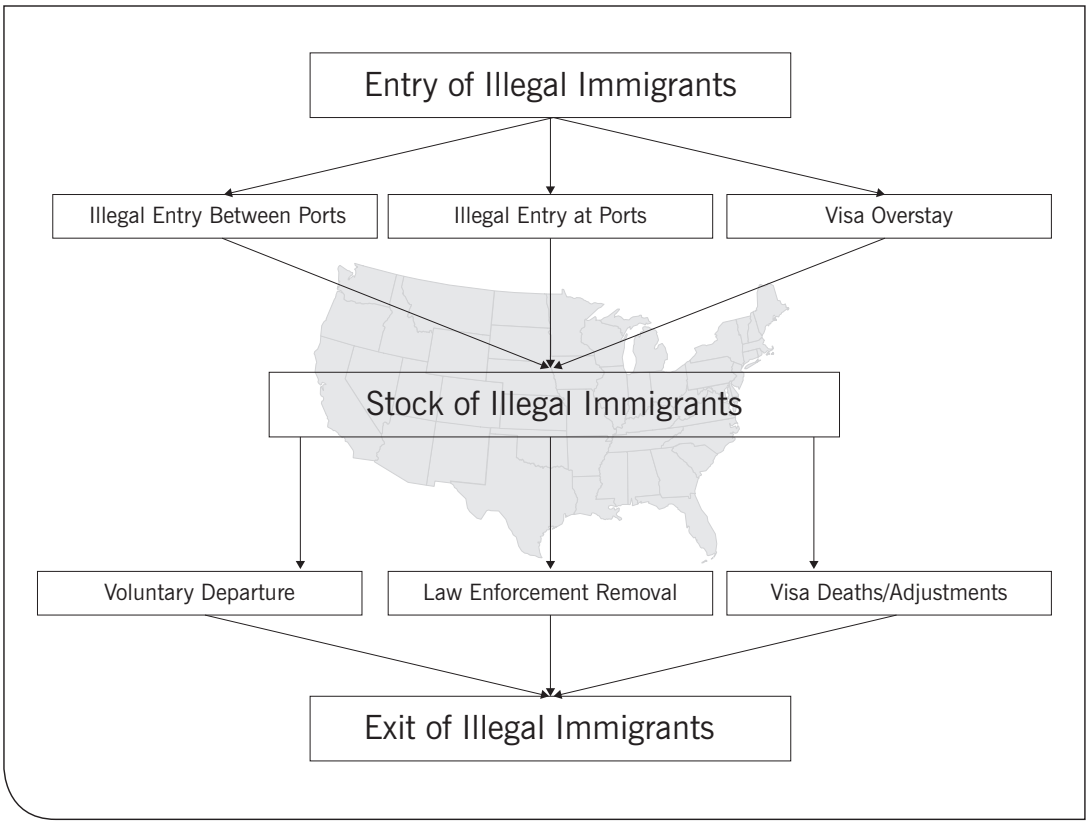
In addition to illegal entry, migrants can enter legally and then move to an illegal status—for example, by overstaying or violating some other condition of their visa. ICE is responsible for law enforcement with respect to visa violations, and interior enforcement more generally. Figure 2 illustrates these violations and their relationships. Figure 2 also includes the exit of illegal immigrants, dividing it into those:

- Who are removed by law enforcement
- Who leave voluntarily
- Whose status is adjusted or who die while in the United States

This understanding of illegal immigration identifies the key outcome performance measures for DHS and its components with respect to this area of law enforcement. Note that the only measures directly observable to the federal government by counting law enforcement administrative records are on exit (law enforcement removal and a portion of deaths and adjustments). All of the other performance measures are based on events unobserved to federal law enforcement officials. Table 1 lists the performance measures identified in Figure 2 and discusses ways in which they might be presented. There are many supporting output, input, and efficiency measures that would also have to be assessed for a comprehensive performance management framework for immigration law enforcement, but the focus here will be on the outcome measures.

DHS currently only measures and reports the last outcome in Table 1, illegal immigrants resident in the United States. Estimates of this outcome have been made since 1980. The primary method of computation is by taking U.S. Census survey data to estimate the foreign-born population in the United States and subtract from it the number of lawful immigrants. In

Figure 2: Illegal Immigration



other words, it combines survey data with administrative records to create an empirical estimate of the illegal immigrant population in the United States.

DHS does not currently systematically estimate or report the other outcomes in Table 1. For illegal entry between Ports of Entry, the largest inflows are generally believed to be the southwest land border (largest inflow) and the maritime border near the Caribbean (second largest inflow)—making these the highest priority for performance measure development. The northern border, other maritime domains, and air domains could be left for future development, illustrating another important principle—performance managers should not try to solve every performance measurement challenge at once; start with the most important areas and then move to lower priority areas after progress has been made on the first tier.

DHS would have several options available for systematic collection of data for these measures, including migrant surveys such as the Mexican Migration Project (MMP)[5] household survey, estimation of known flow (flow that law enforcement observes and either apprehends or does not apprehend), sensor technology, analysis of recidivism data, and red-teaming. Each option has strengths and weaknesses, and there is substantial variation in their costs. The most cost-effective options for the southwest land domain likely are migrant surveys (best way to estimate the apprehension rate, generate data needed by other approaches, and are relatively low-cost) and recidivism analysis (low-cost approach).

For the maritime domain, the problem is different. From 1995 to 2009, the USCG reported the probability of interdiction of illegal immigrants (from which inflow can be computed) who

5. MMP is an academic data collection effort co-led by the University of Guadalajara (Mexico) and Princeton University.

Table 1: Key Outcome Performance Measures for Immigration Law Enforcement

Outcomes	Performance Measures	Comments
Illegal Entry Between Ports	Number of Individuals Entering the Country Illegally Between the Ports of Entry in the Reporting [Year, Quarter, Month]	Four primary variables: • Number of individuals attempting entry • Number of apprehensions • Apprehension rate • Number of successful entries The measure may also be broken out by domain (air, land, and maritime) and by sector within a domain.
Illegal Entry At Ports	Number of Individuals Entering the Country Illegally at the Ports of Entry in the Reporting [Year, Quarter, Month]	Four primary variables: • Number of individuals attempting entry • Number of apprehensions • Apprehension rate • Number of successful entries The measure may also be broken out by domain (air, land, and maritime) and by sector within a domain.
Visa Overstay	Number of Individuals Remaining in the Country Illegally after Violating a Condition of Their Visa in the Reporting [Year, Quarter, Month]	This could be presented in the context of total entries with a legal visa and total departures by legal visa holders. It could also be broken out by visa category.
Illegal Immigrants Resident in United States	Number of Individuals in the Country Illegally in the Reporting [Year, Quarter, Month]*	This could be presented in the context of how it changed from the previous reporting period and the causes of the change, i.e., changes in entry rates versus changes in exit rates. It could also be broken out by location, e.g., state.

** Although not reported in the DHS Annual Performance Report, an estimate of this measure is available from other sources.*

attempted to enter the United States in the maritime domain between the POEs using a known-flow estimate. Known flow may be the most efficient solution in the Caribbean because there are only a small number of source countries (e.g., Haiti, Cuba, and the Dominican Republic), the migrants must follow relatively specific pathways, and, for some countries, such as Cuba, there are special legal provisions that make it in the migrants' interest to declare if they reach the United States. The USCG estimates known flow by using its and other relevant agencies' and governments' data observed through interdiction activities, surveillance, and intelligence.

Unlike inflow between the POEs, where each domain has its own unique characteristics leading to selection of different methods, illegal entry at the POEs is roughly similar across the air, land, and maritime domains. Perhaps the most important difference between domains for estimating outcomes is that CBP usually receives inbound passenger manifests prior to presentation for both the air and maritime domains, but does not for the land domain. Measuring illegal entry at the POEs could be accomplished in a number of ways.

One option would be to expand the existing measurement methods described previously in the Randomized Secondary Screening text box to systematically cover all the potential failure points in the screening process—for example, the accuracy of entry document issuance determinations and the accuracy of screening processes at the POEs themselves. Screening can be divided into physical screening (for example, questioning individuals and use of magnetometers) and information-based screening (for example, collecting biometrics and comparing these

to immigration, criminal, and security databases). Methods for measuring screening accuracy include red-teaming and the already discussed randomized secondary screening following a primary screening encounter.

Finally, the level of immigration that starts out legal but moves to an illegal status (for example, visa overstay) has a few complicating aspects. Perhaps the most common violation is to overstay a visa. For the air and maritime domains, where outbound passenger manifests exist (for commercial air and maritime travel, at least), overstay rates could be estimated by comparing visa records with passenger manifest data. The challenge with this method is that the land domain (pedestrian and vehicle traffic) does not have outbound passenger manifest data. Two options for the land domain would be to use surveys or randomized investigations of expired visas for which land exit is expected (perhaps because entry was by the land domain). For non-overstay forms of compliance (i.e., it is within the period of the visa but the visa holder may have violated another condition of the visa such as refraining from criminal activity), surveys and randomized investigations could again be used to estimate violation rates.

Table 2 summarizes the methods that may be used to estimate the outcomes for illegal immigration law enforcement. These are only meant to be suggestions, as further analysis may be required to finalize the exact approaches to be taken in each case. The goal is to illustrate how the individual detail of specific measurement situations drive the selection of the appropriate methods.

Table 2: Possible Estimation Methods

Outcomes	Estimation Methods
Illegal Entry Between Ports	Southwest Land Domain: Survey Data and Analysis of Administrative Records (Recidivism Analysis) Caribbean Maritime Domain: Administrative Records (Known-Flow Estimates)
Illegal Entry At Ports	Experimental Methods (Red-Teaming) and Inspections (Randomized Secondary Inspections)
Visa Overstay	Visa Overstay Air and Maritime Domains: Analysis on Administrative Records (Government Records Combined with Commercial Records) Visa Overstay Land Domain: Survey Data and Investigations Other Visa Violations: Survey Data and Randomized Investigations
Illegal Immigrants Resident in United States	Analysis of Survey Data and Administrative Records

Implementing the entire program described above would not be easy. Complete cost estimates[6] and implementation plans would have to be developed for each method in each area. Potentially fierce resistance would have to be overcome; some examples of the likely arguments against better performance measurement are provided in the next section. It might take several years before the methods could be fully implemented and sufficiently refined to produce truly reliable estimates. But doing the analysis to develop the strategy will take one major obstacle off the table—the argument that unobserved events cannot be measured and there is no point in trying.

6. The cost estimates should also be put into perspective. Although many investments would have to be made to implement the identified measurement strategy, the cost would likely only be a very small fraction of total expenditure in the mission area and may not be any more than the current expenditure level for all of the proxy measures that DHS currently uses in this mission area. Providing those comparisons could be a valuable way to overcome opposition based on the strategy's cost.

Part V: Challenges and Conclusions

This report has addressed the most direct challenge to measuring law enforcement performance outcomes—how to estimate unobserved events. But even with a sound measurement methodology in hand, there are still obstacles to implementing outcome-oriented data-driven management. In addition, simply measuring the outcomes is not enough—the performance manager must also build an empirical understanding of how law enforcement activities and outside factors affect the outcomes. This section briefly reviews these challenges.

Challenges

There are many challenges facing a performance manager in increasing the use of data to drive decision-making. Five specific challenges are particularly relevant to federal law enforcement. While other challenges will certainly arise in improving the quantitative content of performance management, preparing for these five will assist performance managers in fighting the major battles ahead of them.

Challenge One: Institutional resistance to diverting scarce resources from enforcement to data collection and measurement. Although it may seem obvious that one of the first things a serious manager does when solving a large, complicated problem is to start measuring it, incentives within government organizations are often not aligned to promote this long-run, problem-solving view. In the short-run view often taken by government managers, investing in data collection and measurement takes resources from current enforcement and is thus considered a lower priority (even if it would improve mission performance over the long run).

An example comes from Customs and Border Protection's (CBP) major increase in technology investment along the border during the 2006–2011 time period. CBP greatly increased its use of technology, including sensors and radars to monitor for drug smuggling and illegal border crossings, during that period, but its operating concept was focused on deployment of technology to locations where there were sufficient law enforcement assets, or concurrent increases in law enforcement assets were made, for interdiction. This limited the deployment of the technology and increased its costs. CBP was reluctant to make technology investments where increased situational awareness and data collection would be the primary focus, i.e., where interdiction would not be assured. However, obtaining this situational awareness—i.e., measuring outcomes—could lead to improved allocation of scarce resources and improved outcomes over time, even if there were some "sensor hits" for which no Border Patrol agent was available to respond in the short run. Making this cultural change in a law enforcement organization can be challenging.

The performance manager trying to drive change through a law enforcement organization must predict how this concern will manifest itself in their organization and develop a strategy for overcoming it at the start of the effort. The basic response to the challenge is to explain the

difference between obtaining short-run versus long-run results. The reluctance of operators to make investments for the future is a standard problem and should be countered with a well-reasoned argument for why long-run results will be improved by making investments today.

Challenge Two: Data quality. Although concerns with using an analytic method to estimate an outcome that cannot be directly measured can be exaggerated by opponents of performance-driven management, there are legitimate concerns that the performance manager must take seriously. One set of legitimate concerns is technical; the analytical methods applied will have basic statistical properties such as confidence levels, sensitivity to measurement error, and reliance on assumptions that characterize the accuracy and reliability of the results. The performance manager must explicitly identify and explain these issues to the decision-makers using the data. The goal is to make decisions based on facts and empirical evidence; hiding or minimizing limitations to those facts and empirical evidence harms the decision-making process. It can be tempting for a performance manager to want to minimize these concerns when facing significant organizational opposition to increased data-driven management, but the performance manager must resist this defensiveness and remain transparent and open about the limitations of the data and analysis.

Another set of legitimate issues related to data quality concerns maturing the methods through time. No matter how expert the performance manager is, new methods to solve complicated estimation challenges will take time to get right. Performance managers and decision-makers should be careful not to expect too much too fast from the new attempts to measure their law enforcement challenge. Estimates should be created, socialized, criticized, revised, and improved. Perhaps targets should not be used in the initial years of developing and socializing the data, both because of data quality and maturation concerns and because of the cultural changes that introducing the estimates create.

Challenge Three: Cultural change that comes with quantifying a problem for the first time. Attempting to measure a contentious outcome for the first time is a big step for an organization and exposes it to new areas of criticism and risk. An example comes again from border control. Although DHS does not systematically measure border flow and apprehension rates, there are some estimates available from empirical studies, and these find an apprehension rate between 33 and 50 percent (which implies a flow rate of approximately one million per year) for the southwest land border between the POEs. With some policy-makers arguing that the only acceptable outcome is a sealed border (zero net flow and 100-percent apprehension rate), it is understandable why the less courageous in DHS are reluctant to develop a reportable measure.

But hiding the truth does not change it, and makes it harder both to improve outcomes and have open, fact-based discussions about what outcome levels are acceptable. In fact, the apprehension rate on the border is not that different from the closure rate for many crime categories across the country. For example, the national average closure rate for all violent crimes is about 45 percent and the closure rate for all property crimes is less than 20 percent. When viewed in this context, the DHS border performance may not be considered that bad. Regardless of whether 33–50 percent is a good or bad rate, it is more important that policy-makers cannot have an open debate on whether it is acceptable, whether more resources should be applied to increase it, or whether more efficient methods could increase it at current expenditure levels. The performance manager must be sensitive to how the results will be received and be prepared to discuss why the results, even if they may appear unfavorable at first, need to be produced and presented. Putting them into context as in the above example, with comparisons to violent and property crime closure rates, can be an important way of dealing with this challenge.

Challenge Four: Performance management sits at the intersection of operators and analysts. While the tendency of operators may be to dismiss analysis and data to inform decisions, there may also be a tendency for analysts to over-rely on data and analyses. All data and analyses are imperfect, contain measurement error, and rely on assumptions. Although objective, quantitative performance data are essential to sound management, no linear pathway exists from data to a model to an estimate to a decision. Decisions should be informed by the estimates, but may also be informed by short-run operational realities and constraints, political factors, and stakeholder concerns and interests. Performance measurement analysts must have a seat at the table but must not overstate the usefulness of their results.

Challenge Five: Analyzing the relationship between outcomes, outputs, and inputs. This report is focused primarily on how to measure unobserved events to improve law enforcement, but estimating and reporting crime rates are not the only responsibilities of a performance manager in supporting decision-making. A complete law enforcement performance management framework needs to include empirical analysis of how outcomes, outputs, and inputs are related to each other. It must also include the analytic ability to forecast these relationships into the future.

Similar to the challenges created by the cultural change of measuring the outcomes an organization is trying to impact when they were not previously measured, measuring the direct contribution of the organization's activities on those outcomes also represents a significant cultural change and potential challenge. No program manager wants to learn that the program he or she has been devoted to for years has not had much of a measurable impact on the outcome the program is supposed to effect.

Outcomes are affected by program outputs. Numerous programs are already in place, with other potential programs available, to work toward most outcomes that federal law enforcement organizations are trying to achieve. Selecting among possible programs to achieve desired outcomes most efficiently requires analysis. For example, the DHS component ICE is responsible for interior immigration law enforcement. Some of the programs it uses include worksite investigations, worksite audits, human smuggling investigations, visa compliance investigations, and criminal alien reviews. Like all federal law enforcement organizations, ICE's resources are scarce and it must decide how to allocate those scarce resources across these and other programs—no program receives all the resources it desires. To effectively make these decisions, DHS would need to analyze the effects of these programs and other alternative options on the outcomes it is trying to achieve.

A related analytic change is that the effects of these decisions must be forecast into the future. Setting future targets requires projecting the effects of programs and resources (outputs and inputs) on outcomes. Similarly, adjusting the optimal strategies (i.e., choices among programs) to changing environments requires the ability to project the analyses into the future. The performance manager must thus have the capability to forecast, which frequently requires some capability in modeling and simulation. A model is often constructed using the empirical analysis relating outputs to outcomes based on historic data. This model is then simulated by replacing the historic data with forecasts of the explanatory variables into the future. Although challenging, this type of rigor should be expected in federal law enforcement performance management offices. In the absence of an analytic approach to setting future targets and projecting the effects of strategic decisions, the organization will be left doing so through ad-hoc guesstimation—exposing the organization to dismissal as not serious or knowledgeable about its own mission and operations.

Conclusions

Performance measurement in federal law enforcement is not easy, but it is possible. One advantage that law enforcement organizations have is that the structure for their performance management systems is relatively straightforward. The outcomes they are trying to achieve are the enforcement of the law, and the supporting measures required are the methods by which they enforce the law. Although opponents of data-driven management may argue that the mission of the law enforcement organization is not really to enforce the law, decision-makers and performance managers must remain focused on outcomes and implementing performance-driven management reform in their organizations.

While the structure of their performance management system may be relatively straightforward, a real technical challenge in measuring federal law enforcement outcomes does exist—the outcomes are often unobserved and cannot be directly measured. The primary focus of this report has been on how to measure these unobserved events. A variety of methods were reviewed because there is no one-size-fits-all answer to this performance measurement challenge. Decision-makers and performance managers must systematically analyze their own law enforcement challenges, the available data, and the optimal methods for developing reliable estimates for use in performance management. With this focus in mind, it is possible to bring the radical reforms that have been seen in state and local law enforcement to the federal level and to realize the same impressive improvements in performance that these pioneers have achieved.

References

Giuliani, Rudolph. 2000. "Restoring Accountability to City Government." In *The Business of Government*, edited by Ian Littman, 4–5. Washington, D.C.: IBM Center for The Business of Government.

Goldberg, Jeffrey. 2012. "Why is U.S. Violent Crime Declining?" (Part 2). *Bloomberg* (February 15). www.bloomberg.com/news/2012-02-15/why-is-u-s-violent-crime-down-part-2-commentary-by-jeffrey-goldberg.html.

Morral, Andrew R., Henry H. Willis, and Peter Brownell. 2011. "Measuring Illegal Border Crossing Between Ports of Entry: An Assessment of Four Promising Methods." RAND occasional paper. Santa Monica, California: RAND Corporation.

Smith, Dennis C. with William J. Bratton. 2001. "Performance Management in New York City: *CompStat* and the Revolution in Police Management." In *Quicker Better Cheaper? Managing Performance in American Government*, edited by Dall W. Forsythe, 453–482. Albany, New York: Rockefeller Institute Press.

Acknowledgments

This report was made possible by the contributions of many individuals. Some of them include Ted Alden, John Kamensky, Gary Luethke, Greg Pejic, Mark Phillips, Bryan Roberts, Robert Shea, and Molly Valdes-Dapena. The author would also like to thank the many individuals that assisted within the U.S. Secret Service, the Office of National Drug Control Policy, the Department of the Treasury, and U.S. Citizenship and Immigration Services.

About the Author

John Whitley is a Senior Fellow at the Institute for Defense Analyses (IDA). His work at IDA includes resource allocation and performance issues in national security, defense resource management analysis, and the study of immigration policy. He is also an adjunct lecturer at The George Washington University in the Trachtenberg School of Public Policy and Public Administration where he has taught National Security Economics.

Prior to joining IDA, he was the Director of Program Analysis and Evaluation (PA&E) at the Department of Homeland Security where he led the resource allocation process and the measurement, reporting, and improvement of performance. At DHS, John worked on counterterrorism, immigration, cybersecurity, and disaster management issues. Prior to DHS, John worked in the Department of Defense office of PA&E on defense resource management issues. While at the Department of Defense he also served as a defense fellow in the U.S. Senate in the office of Senator Jon Kyl of Arizona. Prior to returning to Washington, John was a faculty member in the economics department of the University of Adelaide in Australia where he conducted research on agricultural markets and crime. John has also served in the U.S. Army.

John has a PhD and MA in economics from the University of Chicago and undergraduate degrees in Animal Science and Agricultural Economics from Virginia Tech.

Key Contact Information

To contact the author:

John Whitley
Senior Fellow
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311
(703) 575-6615

e-mail: johnewhitley@comcast.net

Reports from IBM Center for The Business of Government

For a full listing of IBM Center publications, visit the Center's website at **www.businessofgovernment.org.**

Recent reports available on the website include:

Assessing the Recovery Act

Recovery Act Transparency: Learning from States' Experience by Francisca M. Rojas
Key Actions That Contribute to Successful Program Implementation: Lessons from the Recovery Act by Richard Callahan, Sandra O. Archibald, Kay A. Sterner, and H. Brinton Milward
Managing Recovery: An Insider's View by G. Edward DeSeve
Virginia's Implementation of the American Recovery and Reinvestment Act: Forging a New Intergovernmental Partnership by Anne Khademian and Sang Choi

Collaborating Across Boundaries

Collaboration Across Boundaries: Insights and Tips from Federal Senior Executives by Rosemary O'Leary and Catherine Gerard
Designing Open Projects: Lessons From Internet Pioneers by David Witzel

Conserving Energy and the Environment

Best Practices for Leading Sustainability Efforts by Jonathan M. Estes
Implementing Sustainability in Federal Agencies: An Early Assessment of President Obama's Executive Order 13514 by Daniel J. Fiorino

Fostering Transparency and Democracy

Assessing Public Participation in an Open Government Era: A Review of Federal Agency Plans by Carolyn J. Lukensmeyer, Joe Goldman, and David Stern
Using Geographic Information Systems to Increase Citizen Engagement by Sukumar Ganapati

Improving Performance

Forging Governmental Change: Lessons from Transformations Led by Robert Gates of DOD and Francis Collins of NIH by W. Henry Lambright
Improving Government Contracting: Lessons from Bid Protests of Department of Defense Source Selections by Steven M. Maser

Managing Finances

Strategies to Cut Costs and Improve Performance by Charles L. Prow, Debra Cammer Hines, and Daniel B. Prieto

Strengthening Cybersecurity

A Best Practices Guide for Mitigating Risk in the Use of Social Media by Alan Oxley
A Best Practices Guide to Information Security by Clay Posey, Tom L. Roberts, and James F. Courtney

Transforming the Workforce

Engaging a Multi-Generational Workforce: Practical Advice for Government Managers by Susan Hannam and Bonni Yordi
Implementing Telework: Lessons Learned from Four Federal Agencies by Scott P. Overmyer

Using Technology

Challenge.gov: Using Competitions and Awards to Spur Innovation by Kevin C. Desouza
Working the Network: A Manager's Guide for Using Twitter in Government by Ines Mergel

IBM Center for
The Business of Government

About the IBM Center for The Business of Government

Through research stipends and events, the IBM Center for The Business of Government stimulates research and facilitates discussion of new approaches to improving the effectiveness of government at the federal, state, local, and international levels.

About IBM Global Business Services

With consultants and professional staff in more than 160 countries globally, IBM Global Business Services is the world's largest consulting services organization. IBM Global Business Services provides clients with business process and industry expertise, a deep understanding of technology solutions that address specific industry issues, and the ability to design, build, and run those solutions in a way that delivers bottom-line value. To learn more visit: ibm.com

For more information:
Daniel J. Chenok
Executive Director
IBM Center for The Business of Government
600 14th Street NW
Second Floor
Washington, DC 20005
202-551-9342
website: www.businessofgovernment.org
e-mail: businessofgovernment@us.ibm.com

Stay connected with the IBM Center on:

or, send us your name and e-mail to receive our newsletters.