**COMPLIANCE AND DETERRENCE RESEARCH PROJECT:
MEASURING COMPLIANCE ASSISTANCE OUTCOMES**

STATE OF SCIENCE AND PRACTICE WHITE PAPER

December 6, 2007

Prepared for: the U.S. Environmental Protection Agency's Office of Enforcement and
Compliance Assurance and Office of Research and Development
By: Dr. Shelley Metzenbaum
with assistance from: Tracy Dyke Redmond, Senior Associate, and Sandra Grund,
Associate, Industrial Economics, Inc. and Adenike Adeyeye, Research Associate
Environmental Law Institute

Acknowledgments:

Disclaimer:

This document is prepared under contract in response to EP07H002309. The views in this document are solely those of the author and do not necessarily represent the opinions of the Environmental Protection Agency or any of its employees.

This report reviews research and information on the current state of science in compliance assistance outcome measurement and develops findings and recommendations based upon technical feasibility, without assessing implementation costs. No work has been done to determine the costs or practicality of implementing the compliance assistance measures discussed in this report. In other words, this report does not determine the costs to the EPA of conducting any specific information gathering, measurement, or statistical analysis nor does this report assess the feasibility of doing so given the OECA's workload, workforce, and budget.

TABLE OF CONTENTS

Assessment of EPA's Current National GPRA Measures Based on the Following Criteria:

A. Comprehensibility and Coherence
B. Motivation Value
C. Diagnostic Value
D. Replication Value
E. Feasibility of Measuring Accurately and Credibly
F. Attribution

EXECUTIVE SUMMARY

The U.S. Environmental Protection Agency (EPA) is committed to using compliance assistance to help regulated entities comply with the environmental requirements that apply to them and improve their environmental performance. EPA uses the term compliance assistance to describe activities, tools, or technical assistance efforts that provide clear and consistent information that helps the regulated community understand and meet its environmental obligations and adopt other environmentally beneficial practices. This paper seeks to inform EPA's efforts to measuring outcomes from CA efforts. Specifically, this research paper responds to the following questions posed by EPA:

1. Is OECA's current Government Performance and Results Act (GPRA) objective the most effective and compelling measure of compliance assistance performance? If not, what are potentially better measures and why?

2. According to the relevant compliance literature, when do regulated entities seek compliance assistance, how and when does compliance assistance drive behavior or motivate change, and how can these outcomes be measured?

3. Recommendations:

- How should OECA act on the answers to questions 1 and 2 above?
- How can OECA use information on the effectiveness of compliance assistance strategically to select and target future compliance and enforcement priorities and activities?

EPA's current GPRA objective for compliance assistance is, by 2008, to prevent non-compliance or reduce environmental risks through EPA compliance assistance by achieving:

- 85 percent of regulated entities improved their understanding of environmental requirements;
- 50 percent of regulated entities improved environmental management practices; and
- 12 percent of regulated entities reduce, treat, or eliminate pollution.

The paper begins by assessing EPA's GPRA measures for compliance assistance against six criteria: 1) comprehensibility and coherence, 2) motivation value, 3) diagnostic value, 4) replication value, 5) feasibility of measuring accurately and credibly, and 6) attribution. These criteria were selected based on discussions with EPA and on a review of the literature and practice regarding the effective use of performance measures. To help assess EPA's measures, the authors conducted a benchmarking analysis of the performance measures other federal agencies use to report on the outcomes of similar efforts, including compliance assistance. We draw insights from those examples in our discussion of EPA's GPRA measures. Performance measures for non-federal agencies (particularly states, but also some sub-national programs of federal agencies) are also reviewed.

Table 1 summarizes the authors' findings and recommendations, based on a benchmarking analysis and a review of the literature, with regard to EPA's current GPRA measures:

1

Table 1: Assessment of EPA's Current GPRA Measures	
Criteria for Review	**Findings**

Comprehensibility and coherence	• Comprehensibility and coherence are essential to realizing the communication power of performance measures
	• Use of a compliance assistance target under GPRA constructively communicates to the public, employees, and Congress that EPA is committed to assisting regulated parties, not just monitoring and enforcing against them.
	• EPA's uses three CA performance measures to assess progress toward both interim and end outcomes the agency seeks to achieve through CA. Conceptually, EPA's measures are comprehensible and coherent.
	• EPA's measures could be improved by:
	    o Eliminating confusion about the magnitude of program impact that arises from measuring percentage change of a population that can vary significantly in number and composition every year.
	    o Providing a context for interpreting the size of the audience EPA is trying to address through CA and the percentage of the audience it reaches.
	• In an ideal world, it would be more coherent for EPA to provide an indicator of compliance for a compliance assistance program, however developing this type of statistic is notoriously difficult and no other federal agencies reviewed uses rates of compliance behavior as a national measure for compliance assistance. Several recent EPA-funded state measurement efforts may lay the needed foundation to measure compliance at the state and national level in a number of programs, especially if EPA continues to support the development and replication of these measurement efforts.
	• In an ideal world, it would be more comprehensible and relevant to the public to include a better measure of the magnitude of pollution, if feasible, rather than solely measuring the percentage of regulated entities that reduce, treat, or eliminate pollution. The experience of other federal agencies suggests EPA might want to explore the feasibility and value of using pollution incidents as a CA (or integrated program) measure.
	• A few federal agencies have adopted national measures for understanding and behavior change, adopting very specific understanding and behavior change targets, noting the target population and the specific concept to be understood or behavioral change to be adopted. In several cases, they pair these national intermediate outcome measures with other end outcome measures. EPA may want to consider being more specific in setting understanding and behavior change targets at the national level, as well as at the sub-national level, linked to high priority problems and opportunities it has identified.

Motivational value	• Performance targets motivate best when they are specific. EPA's CA targets are somewhat specific, but would communicate priorities more concisely and motivate more effectively if they

2

	- were more specific – clarifying the types of understanding to be improved among whom and similarly, priority types of management practices and pollution reducing actions to be adopted by whom. - Performance targets motivate best when they are challenging, but realistic. EPA met all of its targets nationally and in every region last year. To better motivate staff, we suggest that EPA consider making its CA national targets more challenging. However, if targets are so challenging that the agency cannot reach them, the Office of Management and Budget may give the agency a lower score on its Program Assessment Rating Tool (PART) review - Performance measures have a stronger motivational effect when those EPA is seeking to motivate find the measures relevant and useful to their own work, not just as a reporting requirement. To motivate EPA staff with targets and measurement, each discrete CA delivery unit needs to understand what their own targets are and how they are performing relative to their targets. To figure out if CA delivery agents find the measurements useful, the simplest way to figure this out is to ask them. - National measures can be an effective motivator of regulated parties when the measures are broken down by regulatory sector, suggesting national CA sector targets might be beneficial as an industry motivator.
Diagnostic/Learning value	- Measures can be useful diagnostically to illuminate what works – and therefore is worthy of replication -- and what does not – and therefore needs adjustment. Measures can also illuminate problems that need attention and their relative import. Using measures this way supports learning and continuous improvement. - Perhaps surprisingly, national measures have limited diagnostic and learning value. Their real motivation and diagnostic value arise when national measurement is disaggregated to reveal performance variations across the segments. - EPA should experiment with breakout analysis with EPA's current GPRA CA measures to see if they are useful for finding possible problems and other promising practices. - EPA has supported discrete diagnostic studies and experiments in the past to complement its national CA measures. These studies have proven extremely beneficial in advancing measurement methodologies and advancing understanding of the effectiveness of assistance and other compliance assuring efforts. Additional studies would be beneficial, especially to refine understanding of how variations in types of compliance assistance affect intermediate and end outcomes and to further advance measurement capacity. - Many agencies have developed a strong capacity to measure causal factors contributing to problems, which has helped them develop effective prevention strategies. - Continuing the effort to study regional and other sub-national

	(state, local) experiments to look for possible successes worth replicating and learn about problems encountered, building on the preliminary work started with this study, holds promise as a performance-improving approach.
Replication value	• When agencies successfully find more effective and cost-effective practices, replicating them in other locations can improve national performance. 　o Agencies benefit when they systematically search for successful projects on an ongoing basis. 　o Once found, the ability to replicate the project in a second location affirms, although it does not prove, the value of the project approach. 　o If the approach achieves similar performance gains in a second location, it is identified as a good candidate to be considered along with other promising practices for broader adoption. Ultimately, agencies adopt some subset of these practices and promote them nationally. 　o EPA's compliance assistance efforts, informed by the work of Compliance Assistance Advisory Council in 2001, has built a solid foundation that would allow it to embark on a more systematic learning strategy that makes the search for successes and efforts to promote their replication core to the agency's work.
Feasibility	A number of issues can affect the feasibility, accuracy and credibility of EPA's national CA measures. Some of these issues include: • Diagnostic capacity – EPA currently generates CA analytic reports manually from its CA database, greatly limiting the value of CA measures. • Counting the number of assisted entities and understanding how assisted entities relate to the size of the target audience. • The usefulness of the measures to CA delivery agents and reporting variations among them.
Attribution	The EPA compliance assistance program has two distinct attribution issues: one pertaining to other parts of EPA and one to states and localities. Controlled studies are likely to be the best way for EPA to assess the marginal contribution of its assistance efforts compared to other compliance assuring activities. Experience of other federal agencies suggests that attribution issues with states, other levels of government, or suppliers tend to be addressed through the use of more explicit strategies focused on specific problems, sectors, or places.

We also reviewed the published literature to assess what influences regulated entities to seek compliance assistance, how and when compliance assistance drives behavior and motivates change, and how these outcomes can be measured.

Table 2 summarizes the authors' findings and recommendations, based on a review of the literature and experience, about the effectiveness of compliance assistance, factors and circumstances influencing effectiveness, and the reasons regulated parties seek compliance assistance:

Table 1: Assessment of EPA's Current GPRA Measures	
Question	**Findings**
Why reg. parties seek compliance assistance?	Classic deterrence theory clearly explains some of the impetus to seek compliance assistance. Limited anecdotal and quantitative analysis suggest that an increase in inspection levels and enforcement seem correlated with an increase in the number of regulated parties seeking compliance assistance. In addition, some regulated parties reported an ethical sense of responsibility to comply.
How/when does CA drive behavior and motivate change?	Little academic research has been done on the isolated effect of compliance assistance on compliance levels, especially for regulatory programs. Three cases found a likely positive impact of compliance assistance visits on compliance. Only one measured compliance assistance delivered without inspections. Another study with a very small sample tried to compare changes in compliance with and without assistance. This study is useful for suggesting both the possibility of conducting a randomized assignment experiment and the potential value of more focused inquiries on the reasons for non-compliance. Two studies noted changes in intermediate outcomes, specifically increased awareness of regulatory obligations measured by increases in the number of permitted parties after agency notification of specific regulatory obligations. Notifying non-filers via mail seems an effective means to increase the number of registrants/permit holders,[1] even without perceived imminent threat of inspections or enforcement. The lack of understanding of regulatory obligations appears to be a significant barrier to compliance.
Motivating compliance: lessons from social psychology literature	Six factors have been identified that cause one person to say yes to another without first making a utilitarian calculation. These are: reciprocation, consistency, social proof, liking, authority and scarcity. Understanding these principles may prove useful to EPA in designing its compliance assistance efforts.
How type of CA influences effect of CA	The studies and cases we found do not directly compare the effectiveness of one type of compliance assistance against another and no clear conclusions could be drawn about which types of assistance work best when, except that mailings with a threat of follow-up seem to increase awareness of regulatory obligations. EPA should consider doing randomized assignment experiments to better understand which types of assistance work better than others in specific circumstances.

Finally, based on this research, we offer recommendations to EPA in measuring its compliance assistance outcomes. These are described beginning on p.48.

---

[1] CT General Permits

I. INTRODUCTION

The U.S. Environmental Protection Agency (EPA) is committed to using compliance assistance to help regulated entities comply with the environmental requirements that apply to them and improve their environmental performance. Compliance assistance is an essential tool, along with compliance monitoring and enforcement, in encouraging regulated entities to achieve and maintain compliance. Both empirical and theoretical studies find that compliance assistance is a necessary and valued part of the regulatory tool kit. Compliance assistance is often used as the first step in encouraging compliance, to be followed by enforcement actions should regulated entities fail to come into compliance after receiving assistance. This escalating approach to compliance assurance is more cost effective than enforcing against all facilities that are out of compliance, and it is also viewed as a fairer approach than taking enforcement actions against regulated facilities that are acting in good faith, but are not aware of their requirements or how to comply.[2]

EPA uses the term "compliance assistance" (CA) to describe activities, tools, or technical assistance efforts that provide clear and consistent information that helps the regulated community understand and meet its environmental obligations and adopt other environmentally beneficial practices. EPA's Office of Enforcement and Compliance Assurance (OECA) partners with compliance assistance providers to develop and deliver compliance assistance resources such as Web sites, compliance guides, fact sheets and training materials. OECA provides both direct assistance and indirect assistance. To leverage its resources most effectively, EPA directs much of its CA activities to two audiences: 1) small and medium sized businesses and 2) other compliance assistance providers who conduct outreach to such businesses.

Increasingly, OECA is being asked to measure and evaluate the effectiveness of its activities and to improve its results. For example, in preparing the 2007 President's Budget, the Office of Management and Budget (OMB) directed OECA to, "look at using metrics that are statistically valid, including compliance rates, reductions in pollution characterized as to risk, recidivism, contamination incidents, and other tools." OECA is therefore increasing its efforts to identify successful techniques and strategies for promoting, achieving, evaluating, and measuring environmental compliance.

A key challenge in this performance measurement effort is measuring outcomes from CA efforts. In the last few years, EPA Headquarters has undertaken considerable effort to assess the effectiveness of its compliance assistance. For example, the Agency has built a database to collect information from EPA Regions about their CA activities and the results of these activities. It also collects information about the volume of activity at EPA-supported web-based compliance assistance centers, which are run by third-parties through a cooperative agreement. The compliance assistance centers survey site users about changes in their understanding, environmental management practices, and resulting changes in their pollution impact to assess the effectiveness of the websites.

---

[2] Ayres and Braithwaite 1992; Scholz 1984, Coglianese and Kagan 2007, Tyler 1991.

This paper seeks to inform EPA's ongoing performance measurement efforts, and specifically its effort to develop effective and compelling CA measures. Specifically, this research paper responds to the following questions posed by EPA:

1. Is OECA's current Government Performance and Results Act (GPRA) objective the most effective and compelling measure of compliance assistance performance? If not, what are potentially better measures and why?

2. According to the relevant compliance literature, how and when does compliance assistance drive behavior or motivate change and how can these outcomes be measured?
- What are the factors influencing people to seek compliance assistance?
- Under what circumstances does compliance assistance lead regulated entities to affirmatively change their behavior?
- Is there any significant variation in changes in environmental management practices and pollution reduction outcomes based on the type of compliance activity, i.e., between direct and indirect assistance, or between different types of direct or indirect assistance (e.g., direct on-site assistance vs. direct workshops/training)?
- How can the influence of EPA assistance be separated/distinguished from the influence of state-provided assistance and other influences on regulated entities?

3. Recommendations:
- How should OECA act on the answers to questions 1 and 2 above?
- How can OECA use information on the effectiveness of compliance assistance strategically to select and target future compliance and enforcement priorities and activities?

This paper places its primary emphasis on answering the first question and providing recommendations pertaining to the first question. It also provides an overview of what the literature and practice say about the second question.

This paper is based on a literature review conducted by the authors between August and November, 2007. To identify relevant literature, the authors consulted materials provided by EPA (e.g., OECA's 1999 and 2007 Compliance Information Project Literature Summaries and a 2006 OECA report on expanding the use of outcome measures). In addition, we conducted an extensive literature search in multiple fields, including the literature on environmental and regulatory compliance, performance measurement, cognitive psychology, organizational behavior, and social marketing. To identify the relevant literature, we contacted leading experts in each field and conducted web searches. In each area, we looked for well-regarded review(s) of the literature by respected academics and/or a few seminal articles, rather than attempting a comprehensive literature search ourselves.

In addition, this paper is informed by a benchmarking analysis of national measures used by other federal agencies for their compliance assistance programs. We relied heavily on the ExpectMore.gov website for the benchmarking review, looking at agencies that provide compliance assistance, supplemented by the authors' knowledge of other relevant

federal program performance measurement practices. We also looked at compliance measurement experience in the states, much of it supported by EPA funding.

The remainder of this paper is organized around the research questions. The next section assesses EPA's GPRA measures for compliance assistance against six criteria: 1) comprehensibility and coherence, 2) motivation value, 3) diagnostic value, 4) replication value, 5) feasibility of measuring accurately and credibly, and 6) attribution. To help assess EPA's measures, we have examined performance measures that other federal agencies use to report on the outcomes of their efforts, including compliance assistance, and we draw insights from those examples in our discussion of EPA's GPRA measures. We also consider performance measures for non-federal agencies (particularly states, but also localities and another country) to highlight additional insights for how to measure compliance assistance outcomes. Then, we review the published literature to assess how and when compliance assistance drives behavior and motivates change and how these outcomes can be measured. Finally, based on this research, we offer recommendations to EPA in measuring its compliance assistance outcomes.

II. HAS EPA CHOSEN THE MOST EFFECTIVE AND COMPELLING TARGETS AND MEASURES OF COMPLIANCE ASSISTANCE?

The Government Performance and Results Act of 1993 (GPRA) requires that all federal agencies submit to the Director of the Office of Management and Budget (OMB) and to the Congress a strategic plan for program activities. The strategic plans are required to include (among other elements): a comprehensive mission statement and general goals and objectives, including outcome-related goals and objectives, for the major functions and operations of the agency. In addition, each year agencies are required to prepare an annual performance plan covering each program activity set forth in the budget of such agency. The annual performance plans are required to establish performance goals; express such goals in an objective, quantifiable, and measurable form; and establish performance indicators to be used in measuring or assessing the relevant outputs, service levels, and outcomes of each program activity.[3]

In choosing GPRA measures, each federal agency indicates the priority areas in which it hopes to attain significant progress, as well as targets that indicate how far and how fast the agency hopes to make progress toward its priority goals. Each agency is then expected to collect measurements on a regular basis indicating progress toward each specific target, indicating how well, in fact, an agency is advancing its priority goals.

EPA's goals, objectives, and measures related to compliance assistance are shown below in Exhibit 1:

---

[3] Office of Management and Budget, Government Performance Results Act of 1993,
http://www.whitehouse.gov/omb/mgmt-gpra/gplaw2m.html

Exhibit 1: EPA's Current Compliance Assistance GPRA Goal, Objective, and Measures

Goal 5: Compliance and Environmental Stewardship. Improve environmental performance through compliance with environmental requirements, preventing pollution, and promoting environmental stewardship. Protect human health and the environment by encouraging innovation and providing incentives for governments, businesses, and the public that promote

Objective 5.1: Improve Compliance. By 2008, maximize compliance to protect human health and the environment through compliance assistance, compliance incentives, and enforcement by achieving a 5 percent increase in the pounds of pollution reduced, treated, or eliminated, and achieving a 5 percent increase in the number of regulated entities making improvements in environmental management practices.

FY 2005 - 2008 Strategic Plan Sub-objectives	FY 2007 Annual Performance Plan Performance Measures
Sub-objective 5.1.1: Compliance Assistance. By 2008, prevent non-compliance or reduce environmental risks through EPA compliance assistance by achieving: 85 percent of regulated entities *improved their understanding of environmental requirements*; 50 percent of regulated entities *improved environmental management practices; and* 12 percent of regulated entities *reduce, treat, or eliminate pollution.*	**In 2007,** through compliance assistance, EPA will increase the understanding of regulated entities, improve Environmental Management Practices (EMPs), and reduce pollutants. EXTERNAL GPRA MEASURES (TARGETS) **50%** of regulated entities receiving direct CA from EPA reporting that they *improved EMPs* as a result of EPA assistance **15%** of regulated entities receiving direct assistance from EPA reporting that they *reduced, treated, or eliminated pollution* as a result of EPA assistance INTERNAL GPRA MEASURES (TARGETS) **80%** of regulated entities receiving direct CA from EPA reporting that they *increased their understanding* of environmental requirements as a result of EPA assistance **75%** seeking assistance from the CA Clearinghouse or CA centers reporting that they *increased their understanding* of environmental requirements as a result of their use of clearinghouse or centers. **65**% seeking assistance from the CA Clearinghouse or CA centers reporting that they *improved (EMPs)* as a result of their use of clearinghouse or centers. **40**% seeking assistance from the CA Clearinghouse or CA centers reporting that they *reduced, treated, or eliminated pollution* as a result of their use of clearinghouse or centers.
Source: Adapted from "EPA Reporting for CA under the Government Performance Results Act (GPRA)," provided by EPA OECA.	

This paper uses six criteria to consider whether EPA's current GPRA CA measures are effective and compelling:

A) **Comprehensibility** and **coherence**: Do people understand what the EPA CA measures mean and are they logically related?
B) **Motivational value:** Do the measures motivate staff, regulated entities, or suppliers trying to sell help to regulated parties to improve their compliance and environmental practices?
C) **Diagnostic value:** Do the underlying data gathered to measure progress toward the goal identify problems, inform priorities, and help the agency find ways to increase program effectiveness or responsiveness? Do the measures help the agency learn from its own experience to improve?
D) **Replication value:** Does the information gathered help compliance assistance providers find effective practices and stimulate their desire to learn how to replicate the progress made?
E) **Feasibility:** Are the measures accurate and credible?
F) **Attribution**: Does the information gathered enable EPA to assess benefits attributable to CA (when combined with other tools) or to federal investment (when combined with other sources of CA support?

To help assess EPA's measures against these criteria, we have drawn comparisons between EPA's measures and national performance measures used by other federal agencies for both GPRA and the Office of Management and Budget's Program Assessment Rating Tool (PART). Our detailed analysis of other federal agencies measures, and notes on implications for EPA, are included in Appendix A. We also consider findings from non-federal agencies (e.g., states) that are measuring compliance assistance outcomes. We include detailed notes on these measures in Appendix B.

The following discussion assesses EPA's GPRA measures against the six criteria listed above, with insights drawn from other federal and non-federal agencies.

A. Comprehensibility and Coherence

Organizational performance goals are expected to play an important signaling function. Research on private sector organizational performance (non-financial) measures emphasize their importance for communicating organizational priorities, what an organization will do and, by implication, what it will not do to managers, the workforce, and suppliers.[4] Writings on public sector performance measures note a similar public sector value. In addition, public sector measures are also seen as important as a way of communicating organizational priorities and trade-offs to the public and its elected officials, as well as performance relative to those priorities, as a form of public accountability.[5] For goals and measurement to play this important communication and signaling role, measures therefore need to be comprehensible and coherent.

---

[4] Kaplan and Norton
[5] Kettl, KSG

Comprehensibility and coherence are essential to realizing the communication power of GPRA measures. Measures must be articulated so as to make it easy for the public, EPA employees, and its partners (such as states and other agencies delivering compliance assistance) to understand what the measures mean. Comprehensible and coherent measures communicate the Agency's priorities to employees to help them understand where to focus their attention. Comprehensible and coherent measures also support democracy and political accountability by communicating to the public and elected representatives what the agency is trying to accomplish and how well it is meeting its goals.[6] In addition, comprehensible and coherent measures help agencies enlist external allies such as other levels of government, private sector consultants, suppliers, and local not-for-profit organizations.[7] Finally, these measures inform regulated parties and those receiving government services about government priorities.

How well do EPA's current GPRA compliance assistance objectives and measures meet the coherence and comprehensibility test?

EPA's current GPRA CA measures have several strengths, as well as some opportunities for improvement. First, and perhaps most important, just having GPRA compliance assistance targets communicates to the public, employees, and Congress that EPA is committed to assisting regulated parties and that EPA does not plan to rely solely on monitoring and enforcement tools to achieve compliance and environmental gains. Second, it signals EPA's commitment to measure and improve the effectiveness of its compliance assistance efforts. This is an important message to send, especially since the availability of assistance affects public and regulated party perceptions of agency fairness[8] and since EPA Regions often feel confused about how much emphasis to place on enforcement and how much on assistance.[9] EPA's CA measures effectively communicate the message that compliance assistance is an important EPA strategy.

EPA's three CA performance measures assess progress toward both interim and end outcomes the agency is seeking to achieve through CA. Thanks to the rapidly emerging science of cognitive psychology, government agencies have increasingly begun to appreciate that people move through multiple cognitive stages on the path to the adoption of regulatory and beyond-compliance practices: attention/awareness, comprehension, change in attitudes/beliefs, motivation to change behavior, and actual behavior change.[10] Many agencies have incorporated these stages into program planning, and measure progress at certain stages as intermediate outcome indicators. A CA logic model might classify initial outcomes as changes in awareness, attitudes, understanding, knowledge, and skills resulting from program outputs. Intermediate outcomes involve changes in behavior that are broader in scope than initial outcomes, and often build upon the progress achieved. End outcomes parallel the overarching goals of the program and are the environmental improvements and public health benefits that flow from the behavioral,
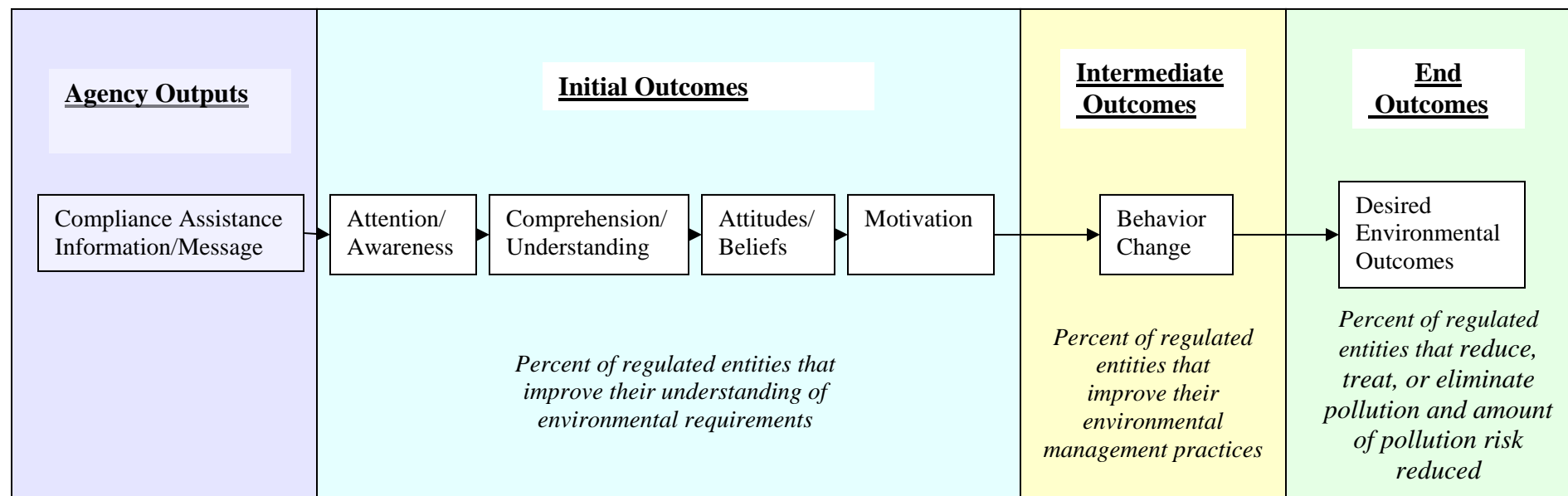
---

[6] GPRA 1993, GASB, KSG.

[7] KSG

[8] Tyler

[9] Author's observation as former EPA Associate Administrator for Regional Operations.

[10] Wogalter

procedural, and operational changes. EPA CA can be seen as having logically adopted one performance measure for the initial outcome stage, one intermediate outcome measure, and one near-end outcome measure as shown in Exhibit 2 below.

Exhibit 2:
ALIGNMENT OF PERFORMANCE MEASURES WITH
COMPLIANCE ASSISTANCE LOGIC MODEL*

Agency Outputs	Initial Outcomes	Intermediate Outcomes	End Outcomes
Compliance Assistance Information/Message	Attention/ Awareness → Comprehension/ Understanding → Attitudes/ Beliefs → Motivation	Behavior Change	Desired Environmental Outcomes
	Percent of regulated entities that improve their understanding of environmental requirements	*Percent of regulated entities that improve their environmental management practices*	*Percent of regulated entities that reduce, treat, or eliminate pollution and amount of pollution risk reduced*

* Performance measures shown in italics. Note that abbreviated chart does not show a complete logic model for EPA's CA activities. Moreover, there are loop-backs in this model. Attitudes and beliefs, for example, can increase attention and behavior change can influence attitudes.

The "end" outcome EPA uses to measure its CA objectives, an increased percentage of regulated entities reducing, treating, or eliminating pollution, addresses an issue of great concern to the public and the agency. The dangers of pollution and the need to reduce or handle it properly are well understood. Signaling the agency's intent for an additional 12 percent of regulated entities to reduce or improve their handling of pollution is likely to make sense to the public in the context of the agency's mission and public concerns.

It would be even more comprehensible and relevant to the public to include an end outcome measure of the magnitude of pollution reduced and properly handled, rather than solely measuring the percentage of regulated entities that reduce, treat, or eliminate pollution. The experience of other federal agencies suggests that it may be useful to measure another aspect of the program's outcomes by measuring the reduction in the number of pollution *incidents*. For example, the U.S. Coast Guard Marine Environmental Protection (MEP) program, the DOT Federal Motor Carrier Safety Administration Operations[11] and Programs office, and the DOT Pipeline Safety Administration[12] all count the number of pollution incidents as one of their national program measures. (The Coast Guard MEP program also counts the volume of pollution spilled, although it no longer uses this as a national GPRA measure.) Indeed, a large number of the U.S. federal regulatory agencies we reviewed for this analysis count specific types of unwanted events (e.g., the number of oil and chemical spills, workplace fatalities, etc.) as one of their GPRA indicators related to compliance assistance. Most agencies (OSHA, MSHA, NHTSA, Coast Guard) attempt to count all events of the specified type (e.g., OSHA uses a census of all workplace fatalities), although OSHA also uses statistical sampling techniques to measure workplace illness and injury incidents.

It is also worth noting that the federal regulatory agencies counting unwanted incidents combine measurement of all of their compliance assurance activities. They do not try to measure the effect of assistance efforts discretely. Instead, they use a single, outcome-focused measure of unwanted incidents for all compliance assurance work, opening up a broader range of program delivery options, affording greater discretion to field staff to adapt the action to the situation while keeping the pressure on them to be effective.[13]

Another of EPA's GPRA CA objectives is increasing the percentage of regulated entities improving their environmental management practices (EMP). This measure addresses an important intermediate outcome, since one way to achieve pollution reduction and environmental improvements is to get regulated parties to adopt mandated and effective voluntary environmental management practices. Conceptually, this indicator meets the coherence test. Because improving EMP is a relatively general concept, however, it is unlikely to communicate the Agency's priorities about the types of EMP of particular concern. This is especially true since there are a diverse array of environmental management practices, not all of which have beneficial impacts on pollution and risk. In our review of performance measures that other federal agencies use, we found several behavior change measures all of which were more specific than EPA'S CA measure.

---

[11] http://www.whitehouse.gov/omb/expectmore/detail/10002250.2004.html
[12] http://www.whitehouse.gov/omb/expectmore/detail/10002264.2004.html
[13] Kowaleweski

They identify the types of behavioral practices that the agency is encouraging, and in some cases the measures identify a specific audience whose behavior the agency hopes to change. For example, NHTSA adopted a GPRA measure for safety belt use and use of child restraints, while FDA adopted a measure to reduce medication errors in hospitals by increasing their adoption of bar code medication administration technology. These examples suggest that the communication value of EPA's behavioral measure could be greatly enhanced by specifying the EMP performance measure more narrowly to indicate what types of behavior change are sought by whom (target audience).

Given that one of EPA's primary goals in delivering compliance assistance is to prevent non-compliance and reduce environmental risks, including an explicit measure of compliance behavior would also enhance the coherence of EPA's GPRA CA EMP measure (at least conceptually). The EMP measure includes compliance practices among all the practices it measures, but it is not possible to distinguish the percentage adopting compliant practices from the percentage adopting voluntary practices. However, in practices, measuring compliance is notoriously difficult.[14] EPA previously experimented with a statistically valid method for measuring compliance and learned an important lesson: it proved dauntingly expensive, relative to program budget, because the large number of variations in the types of facilities regulated and the numbers of regulatory obligations required a huge sample size. In our review of the practices of other federal agencies, we did not find any that use compliance as a national CA measure. The IRS does measure national compliance levels for individuals and for business periodically, using statistically valid random sampling, but does not use compliance as a national GPRA measure. Instead, it uses periodic compliance evaluations to inform its planning and measures the quality of its customer interactions as a national GPRA measure for its assistance efforts. State governments have begun to develop some promising compliance measurement practices, many with financial support from EPA, and these may prove helpful to EPA in developing its national CA measure. This are described in Appendix B and discussed further in the Recommendations section.

EPA's third objective deals with an initial outcome, namely, understanding regulatory obligations. Measuring "understanding" is conceptually coherent with EPA's goals and its other two measures. As with the EMP objective, the focusing function of EPA's "understanding" objective is hampered by its lack of specificity. As with any performance measure, greater specificity communicates the agency's emphasis more clearly to the public and better focuses employee efforts. Other federal agencies we found

---

[14] Regulated entities have multiple compliance obligations, so EPA would need to decide what constitutes non-compliance for the purpose of the GPRA measure. For example, is a facility non-compliant if it fails on just one compliance obligation, or does it need to exceed a certain threshold? Moreover, EPA is not able to monitor all regulated entities to monitor compliance, and so non-compliance rates are a direct function of the compliance monitoring rate, and how the agency selects which entities it will monitor (e.g., by targeting entities suspected of non-compliance, or by monitoring a sample of all regulated entities). Further, compliance monitors tend to use their discretion about where to focus their inspections, so compliance rates may not be meaningful unless the Agency puts in place a standardized inspection checklist that all inspectors use to monitor compliance. Finally, different inspectors clearly vary in their decisions about whether or not similar problems deserve a non-compliant determination.

that use GPRA "understanding" measures specify the type of "understanding" they seek to increase. FDA, for example, uses the following measures for "understanding": the percentage of American consumers who correctly identify that trans fat increases the risk of heart disease; the percentage of consumers who can correctly identify that saturated fat increases the risk of heart disease; and the percentage of American consumers who correctly identify that omega-3 fat is a possible factor in reducing the risk of heart disease. The communication value of EPA's understanding measure would be greatly enhanced with greater specificity about whose understanding of what is to be increased.

EPA has translated its three Strategic Plan Sub-objectives (shown in the left-hand column in Exhibit 1) to three Annual Performance Plan Performance Measures (shown in the right-hand column in Exhibit 1). These annual performance measures, unlike the objectives, assess changes only among those receiving CA. Achieving change among only those receiving assistance is less comprehensible than EPA's stated objectives, especially without better information about what percentage of the intended audience is actually receiving assistance. Without knowing what portion of the target universe EPA is reaching, information about what portion of those receiving CA is much less meaningful. Changes in the number or percentage of entities assisted may greatly affect both EPA's impact and the interpretability of the measures. For example, did EPA assist 10 entities or 1 million? If EPA assisted only 10 entities, EPA could achieve its target that 15 percent of regulated entities receiving direct assistance from EPA reduce, treat, or eliminate pollution as a result of EPA assistance if only one regulated entity reduced, treated or eliminated its pollution. On the other hand, if EPA provided direct assistance to one million regulated entities, 150,000 would have to change the way they handle their pollution in order to meet this goal.

The absence of information about how many regulated entities EPA is reaching also makes trend comparisons nonsensical if the number or nature of those assisted changes significantly from year to year. Consider the following possibility. In year 1, 85% of 1 million regulated parties, or 850,000, whom EPA helped with CA improved their environmental management practices. In year 2, 65% of the 2 million regulated parties EPA assisted, or 1.3 million, reported that they improved their EMP following EPA CA. EPA clearly had a greater impact in Year 2 than in year 1, helping 1.3 million rather than 850,000 regulated parties. Yet with EPA's current measures, it would appear to have performed more poorly in year 2: only 65 percent of those helped would report improvements compared to 85 percent the previous year. Similarly, if EPA focused CA on very different regulated parties each year, changes in the measures from year to year would not be relevant. This issue of a shifting denominator in the ratio of entities that change their behavior compared to entities that EPA assists seriously limits the interpretability of its GPRA measures.

Performance measurement approaches from other federal agencies suggest ways that EPA could enhance the coherence and comprehensibility of its performance measures in regard to the issue of the number of entities that EPA is reaching through CA. For example, the Substance Abuse and Mental Health agency, which similarly uses self-reported measures of change among those assisted, complements its measures of

16

percentage of clients reporting change with another GPRA national measure, the number assisted. Clarifying the number receiving assistance would enhance EPA's performance measure. EPA already counts and could report the number of entities receiving direct assistance, but the agency would be well served by devising a better method to estimate the number of entities using web-based Compliance Assistance Centers. The current measure, user sessions, is not an adequate indicator of the number of Center users, as explained below in the feasibility discussion.

A related and significant coherence problem with EPA's current CA measures, especially for the tax-paying public and their agents (the President, OMB, and Congress), is the difficulty of understanding the magnitude and nature of the problem EPA is trying to address (how big the non-compliance problem is that EPA seeks to address through compliance assistance) and what portion of it EPA is trying to address. EPA has clarified this to some extent by its emphasis on small business and its establishment of 13 sector-focused CA centers. Providing better information about the size of the population needing assistance (how big is the regulated community, what percentage need assistance, and how much of that community EPA is able to reach) would further strengthen public understanding of EPA's performance measures. For example, if EPA is reaching the vast majority of those needing assistance, then a 85 percent improvement in understanding, 50 percent improvement in environmental management practices, and 12 percent reduction in or improved handling of pollution (EPA's current targets) is readily understandable. On the other hand, if EPA is only able to reach a small minority of the population with compliance assistance each year, it would take EPA many years just to reach the entire target population, and even more time to improve its understanding, practices, and pollution impact. Thus the comprehensibility of EPA's CA annual measure would be greatly enhanced by providing information about the size of the universe that EPA is trying to reach, not as a CA measure but to provide a context for understanding the import of the measures EPA is reporting and to inform resource allocation decisions.

B. Motivation Value

Well-framed and well-used measures have the power not only to communicate, but also to motivate. Targets and measurement can motivate EPA, partner CA delivery agents and even regulated parties to work harder and smarter. There are several characteristics of performance measures that work well to motivate EPA, its partners, and regulated entities. These characteristics include measures that are:

- Specific, challenging, and clear;
- Relevant and used;
- Comparable to the past and to peers;
- Motivating to regulated parties; and
- Motivating to grantees

The following sections discuss EPA's current performance measures in light of these characteristics.

Specific, Challenging, Clear. Cognitive psychologists have found that when targets are specific and challenging but not overly ambitious or complicated, they focus, energize,

17

and encourage persistence among individuals and organizations. People instinctively like to do well, and measurement enables them to see how they are doing and adjust their actions accordingly. It enables them to work harder and smarter to achieve specific goals even without the explicit promise of reward or threat of punishment.[15]

Targets and measurements are more motivating when they are specific and not overly complicated. EPA's CA target is already somewhat specific, stating how much change (85/50/12 percent) it wants by when. Conceptually, this target has the potential to send a motivating signal, but the measures are complicated by the possibility that the nature and number of regulated entities receiving assistance changes significantly each year. Consider, for example, if EPA delivered a high percentage of one-on-one assistance one year and a low percentage the next. A quick look at the ISIS data suggests EPA would score lower on pollution changes the second year because those receiving on-site assistance make pollution adjustments and record changes more rapidly than those attending a workshop. To make its current targets more effective motivators, EPA needs to fix the "denominator problem," by clarifying the nature and number of regulated entities receiving assistance each year.

To motivate, targets also need to be challenging. (OMB's PART embraces this scientific finding with criteria 2.2., which scores a program lower for insufficiently ambitious targets.) EPA met all of its targets nationally and in every region last year. There are many ways the agency could do this. For example, EPA could identify a low-performing or high-risk area and target it for improvement. The FAA did that by adopting one national target to reduce the number of air carrier accidents in Alaska and a second national target to reduce the number of general aviation and nonscheduled Part 135 fatal accidents, in addition to a more general target of reducing the total number of accidents per departure. Or, EPA could adopt a new target to increase the number of parties assisted or the percentage of the universe reached, in addition to improving understanding, EMP, and pollution-handling practices among them. A third way to make the targets more challenging would be to switch the current CA method for measuring the pollution effect from a binary (Y/N) indicator to an indicator of the quantity of pollution reduced or frequency of pollution incidents.

Based on our review of the experience of other federal agencies, we offer two cautions with regard to setting challenging targets for a federal program. First, under PART criteria 4.2, OMB will reduce a program's score if it misses its targets. Some OMB examiners give partial credit even when targets are missed, provided a program is making progress toward its targets. They tend not to give full credit for missed targets, however, causing some agencies to adopt timid targets at the expense of adopting challenging ones that tend to be more motivating. Second, a missed target also provides members of Congress an easy opportunity to criticize a federal agency in order to attract media attention. NHTSA, which initially set extremely ambitious targets, came under fire by both OMB and a member of Congress when it missed that target. Over time, NHTSA has adopted less ambitious, more realistic, targets than it originally set.

---

[15] Locke and Latham

Relevant and Used. Performance measures have a stronger motivational effect when those EPA is seeking to motivate find the measures relevant and useful to their own work, not just as a reporting requirement they need to meet. In the case of EPA's CA work, those to be motivated include CA managers, EPA regions, and CA centers. It might also include states and localities delivering CA funded by EPA. A key question EPA should consider is whether or not its CA measures, as currently framed, are useful to CA providers. The simplest way to figure this out is to ask them. In addition, EPA CA management could ask CA providers about the types of analysis they would find useful.

To motivate EPA staff with targets and measurement, each discrete CA delivery unit needs to understand what their own targets are and how they are performing relative to their targets.[16] For example, are CA providers reporting directly on the national targets or on a distinct, contributory sub-target (e.g. developing a CA training program for small farms that will be used nationally)? This characteristic is sometimes referred to as the need to have measures that "cascade down and roll back up" EPA's decision to measure each region and CA center sends the signal that each delivery unit is expected to meet the national targets.

Comparable to the Past and to Peers. National GPRA measures can have motivational power if they support comparisons to the past and fair comparisons to peers. Most people (and organizations) like to do better than they have done in the past. As already discussed, EPA's current GPRA CA measures need to be adjusted to tap the motivating value of trend comparison. With regard to peer comparisons, we did not identify any other countries using national measures similar to the ones EPA is using for CA that might be useful for and motivating as a peer comparison. We did identify an indicator used by another country that might be useful for EPA's CA program and ultimately as a possible peer benchmark: the United Kingdom's system for counting the number of serious pollution incidents nationally and by sector every year.

Motivating to Regulated Parties. National measures can also be an effective motivator of regulated parties when the measures are broken down by regulatory sector. For example, when the U.S. Coast Guard shifted to outcome-focused goals, it chose fatality rates as a national safety measure and then compared fatality rates for different types of maritime vessels. This analysis identified towboats as the vessel type with the highest fatality rate. This fact, initially resisted but eventually accepted by the towboat industry, enabled the Coast Guard to enter into a very productive assistance-oriented long-term relationship with the industry. The fatality rate was cut by half in one year and the Coast Guard and tow boat industry continue to meet quarterly to study recent experience and identify areas for further improvement. After it succeeded with the tow boat industry, the Coast Guard started to develop a similar assistance-oriented effort focused on fishing vessels, which proved to be the most dangerous once the tow boat industry fatality rate fell.

EPA's current GPRA CA measures have little meaning to regulated parties in the aggregate, but it is not hard to imagine how adjustments could be made, making them sector specific, that would give them more meaning and be more motivational for

---

[16] Kaplan and Norton

regulated parties. Industry might be very interested in, and motivated, for example, by information about awareness of regulatory obligations, understanding of key concepts, adoption of specific practices, and pollution reduced for specific regulatory sectors.

Motivating to Grantees. EPA delivers a significant amount of CA through 14 web-based CA centers each of which serves a different constituency. In principle, EPA's national CA measures could and should motivate them to do better, especially if EPA and the Centers can figure out how to count the number of regulated entities using the Center sites (not the same as user sessions) and if EPA adopts more specific and challenging targets.

C. Diagnostic/Learning Value

Measures can be useful diagnostically to illuminate what works – and therefore is worthy of replication -- and what does not – and therefore needs adjustment. Measures can also illuminate problems that need attention and their relative import. This supports learning – a program's ability to learn from its and others' experience and continually improve. This section considers the illumination, or diagnostic, value of EPA's current GPRA CA national measures.

Measurement helps an agency or program:
- Identify societal and program problems that need attention;
- Inform priority-setting;
- Assess if a tool, project, and program is working and how it can be improved;
- Identify more effective intervention approaches that deserve replication;
- Identify less effective interventions that need improvement (including replacing it with a more effective approach);
- Understand causal factors contributing to problems and to progress;
- Support or challenge cause/effect relationships and treatment/effect hypotheses:
- Track progress toward the outcome and where progress may get stopped;
- Identify barriers impeding progress toward the outcome;
- Improve understanding of how to influence the target audience; and
- Identify measurement inconsistencies.

Perhaps surprisingly, national measures have limited diagnostic value. National measures powerfully communicate agency/program priorities and progress. They can also motivate through trend analysis and on the rare occasions when benchmarking with other nations is possible, as discussed in the preceding section. But their real motivation power, as well as their diagnostic value, arises when national measurement is disaggregated by segment (e.g., performance unit such as the region or CA center, type of facility helped, or type of assistance provided) that reveal performance variations across the segments.

There are three exceptions to this general rule that disaggregated measures are more helpful for diagnosis than national measures. One exception is when the characteristics of program assistance and those being assisted are highly similar. In that situation, national measurement can provide diagnostic information by letting an agency know whether or

not it should continue what it is already doing. Second, when an agency measures each intermediate outcome in the outcome-sequence (logic model), national measures indicate the stage at which progress has stopped. This helps an agency pinpoint where intervention adjustments are needed. Third, the individual outcome data points that roll up to a national measure are useful as the dependent variable in regression analyses to sort out how much outcome change is attributable to agency action and how much to other factors influencing outcomes. Over the long term, the ability to conduct regression analyses is a very important use of the data used to construct national measures, especially since it supports independent third-party program evaluations. But regression analyses are costly to do on an ongoing basis because they require the accumulation of a large amount of data for the other possible explanatory (independent) variables (including compliance assistance) that might account for program performance. (It is noteworthy that in a background literature search for this paper, we found no regression analyses that studied compliance assistance as a discrete variable. This may be due to the fact that EPA and other regulatory agencies just started trying to track compliance assistance.)

Generally speaking, though, to tap the diagnostics power of measures, agencies segment or disaggregate the national measures they collect. The federal agencies we studied that use measurement most effectively for diagnostic purposes conduct extensive breakout (disaggregated) analyses, pay careful attention to causal factors to understand the cause/effect relationship better and to assess the relative import of each causal factor, conduct discrete measured experiments to assess the treatment/effect of discrete tools or campaigns, and conduct supplementary investigative studies to understand barriers to effective program implementation, characteristics of the target audience likely to influence their responses, and other missing pieces of the picture. These approaches are described briefly below.

Agencies conduct breakout analyses to find performance variations among different subsets, sorting by characteristics associated with performance variation. For example, the Coast Guard Marine Environmental Protection Program (MEP) shows how agencies can report national measures, but then disaggregates these measures for diagnostic purposes. Coast Guard's MEP uses the five-year average number of chemical discharge incidents and oil spills greater than 100 gallons per 100 million tons shipped as a national measure for PART. This measure has little diagnostic value in and of itself. However, the agency also tracks disaggregated data (e.g., number and volume of spills by specific waterbody, location, source, type of oil, and Coast Guard District). These more detailed data are of great value in diagnosing problems and pinpointing what strategies are working. For example, Coast Guard data show a dramatic drop in the volume of oil discharged from large spills in 1991, and another substantial drop in 1997. Correlating these changes with changes in Coast Guard strategies or other external events (such as passage of the Oil Pollution Act of 1990) could be very helpful in understanding what strategies are working. Examples from other agencies are provided in Appendix A.

Performance management experts offer a catalogue of characteristics to use for breakout analyses, including the organizational unit delivering the service, characteristics of the

21

parties served, geographic location, and type of treatment.[17] Agencies use these breakouts to look for:

- Previously unknown problems;
- Subsets of the measured universe with lower performance that are candidates for priority attention; and
- "Positive deviants" (clusters that outperform other clusters or, within a cluster, individuals that outperform others in the same cluster.)

Many agencies also count or characterize causal factors, both negative and positive, known to affect outcomes. Disaggregating performance analyses by causal factor reveals each factor's relative import. Two examples are the National Highway Traffic Safety Administration and the U.S. Coast Guard. NHTSA has developed a tool it calls the Haddon matrix to structure its causal analysis. The Haddon matrix is a 3 x 3 table filled in by an on-site observer. Column headings ask observers to note key characteristics before, during, and after every fatal accident. Row headings ask observers to record conditions in three key categories: operator, equipment, and environmental (conditions of the situation, such as the type of intersection).[18] The U.S. Coast Guard similarly looks at conditions and causal factors associated with spills and vessel fatalities. It notes the "who, what, when, where, and why" of each unwanted incident. For spills, for example, it notes the time of a spill, operations when the spill occurred, location, source, and volume.[19]

Breakout analyses reveal problems agencies might not previously have known about. For example, sorting the number of spills by time period revealed to a regional office of the U.S. Coast Guard that a high percentage of spills occurred at night. Breaking out fatal accident data by type of vehicle showed NHTSA that the rise in motorcycle fatalities was overwhelming the steady decline in automobile fatalities, causing the national fatality rate to rise.

Breakout analyses also help agencies identify natural experiments taking place around the country (and world.) When Kentucky and Louisiana changed their motorcycle helmet laws and no other states did, a NHTSA-funded researcher compared changes in fatality and injury rates in the two states with changes in all other states to calculate the human benefit of strong motorcycle helmet laws.[20] NHTSA also scans for anomalies, unexpected performance variations, to find natural experiments in the states.[21] This led it to the discovery of the enormous fatality-reducing effect of California's primary enforcement law that allows police officers to stop automobiles to check for belt use. (Other states with belt laws only allowed officers to check belt use when pulling cars over for other reasons.)

---

[17] Hatry
[18] Runyon
[19] Stalfort and McHenry
[20] Metzenbaum 2006
[21] Metzenbaum 2005

A number of federal agencies (HUD, D.C. child support, Education) are increasingly designing experiments to assess whether or not a specific type of tool, project, campaign, or program is effective before rolling it out in multiple locations. Agencies randomly assign those who want to be assisted to experimental and control conditions, then measure how different groups are affected.[22] A few federal agencies have enlisted states to participate in random assignment experiments, randomly assigning states to an experimental or control group. Early fears about the unfairness of providing service to one group and not another seem to have been put aside with the realization that programs unable to provide assistance to everyone are already making choices about who gets assisted and who does not, and an acknowledgment that, since the effects of assistance are unknown, it is not unfair to assist one group without assisting everyone.

EPA currently conducts some breakout analysis of its national CA measures, looking at performance variations by region and by CA Center. A quick look at three years of analysis done for the 14 Centers suggests two possible "positive deviants" worthy of further study. Of the centers with thirty or more respondents to the annual survey, the Automotive Recycler center consistently scored above the national average on all three CA measures. The Printers' Assistance Center scored above average on all but one of nine scores. Is there something these two centers are doing that others are not and that the other centers could do to improve their performance? The breakout analysis suggests this as an area for further investigation. In addition, the use of the measures to assess a hospital-focused initiative of EPA Region 2 hints at the potential of breaking out measurement by sectors. (See Appendix B.)

EPA might be able to apply useful breakout analysis to EPA's current GPRA CA measures, even with the problems noted above, to find possible problems and other promising practices. For example, EPA might try to answer the following questions:

- Do all workshops have similar performance on all three CA measures or do some score much higher or lower than others? If so, why?

- Is all on-site assistance of similar effectiveness or do some score much higher or lower than others? If so, why?

- How does the performance of the different types of assistance compare? How does the number assisted and the number providing assistance compare for different types of assistance?

- How does performance compare in different sectors assisted through direct assistance? If there is variation, is any of it explained by the CA provider?

- If information about the number of regulated parties in different sectors in each region can be found, how does the percentage of regulated parties in each sector vary by region? If one region clearly specializes in a sector others are not addressing, is its performance good and if so, is it feasible for that region to provide assistance to the whole sector?

---

[22] Ayres, Lewin

23

- What types of behavior change were the most commonly reported and how does that vary by sector/delivery agent? What types of pollution changes were the most commonly reported and were there variations among sectors/delivery agents?

- What explains the significant differences in entities reached and outcome variations across regions? Is it attributable to the problems they are addressing, the characteristics of the regulated entities in the region, or the way they are reporting?

  Such breakout analyses would reveal additional insights about the strengths and limits of EPA's current GPRA CA measures, especially if they provide useful insight to CA delivery agents or provoke additional questions.

EPA has already funded a number of diagnostic studies as a complement to its national CA measures to inform strategies and enhance the comprehensibility of the measures. For example, EPA conducted a study of the awareness of two of the CA centers, to assess the percentage of the regulated sector served by the Center aware of each Center's existence and found a relatively low level of awareness. A follow-up study would inform EPA if awareness of the Centers and the assistance they provide has increased. The U.S. Department of Labor has since conducted a similar study, using a slightly different methodology, to assess awareness and use of its assistance.[23] EPA has also funded a number of studies that strongly suggest the value of compliance assistance, especially when paired with monitoring and enforcement threats. (See Appendix B.)

In addition, if any EPA CA delivery agents are using pre- and post-assistance surveys/tests, closer attention to the details (answer by answer) of the tests may reveal specific aspects of assistance that work and specific aspects that did not (e.g., what types of practices changed after a workshop? Did the changes correspond to the content of the workshop?).

D. Replication Value
The previous section discussed the need for diagnostic measurement and analysis to identify effective tools, projects, and interventions. When agencies successfully find more effective and cost-effective practices, replicating them in other locations can improve national performance. As noted earlier, disaggregated national measures support the discovery of positive deviants worthy of replication. So can experiments testing the effectiveness of specific assistance efforts.

When promising practices are identified, disaggregated data can also be helpful in tracking the effectiveness of replication efforts, provided changes in outcomes can be tracked by location as replication efforts are introduced. For example, in a hypothetical example, if EPA starts to segment its analysis and notices that assisted colleges in region Z report a much higher level of change in understanding regulatory obligations, key EMPs, and pollution reductions than assisted colleges in other regions, it might take the following steps to improve national performance.

---

[23] Lotspeich and Fix

- First, it would want to confirm that Region Z's CA approach is in fact more effective than that of other regions and not primarily attributable to differences in measurement and reporting practices.

- Next, it would want to test the replicability of Region Z's successful approach in a second location. If the second location does not experience changes similar to those seen in the first, it should naturally raise questions about the value of repeating the approach elsewhere.

- If the approach achieves similar performance gains in the second region, EPA would treat it as a good candidate for broader adoption to be considered along with other promising practices successfully replicated. EPA's national measures, disaggregated, would, it is hoped, eventually support tracking of replication efforts.

E. Feasibility of Measuring Accurately and Credibly

EPA currently uses a combination of surveys, on-site visits and re-visits, self-reporting, and checks of the compliance database to compile its national CA measures. This section briefly lists some of the issues that can affect the feasibility, accuracy and credibility of EPA's national CA measures. In considering the feasibility and accuracy question, it is important to keep in mind that measures do not need to be 100 percent accurate to be credible and useful. As Former U.S. Postmaster General Marv Runyon once said, "I will tolerate a 40% error rate if I can learn from the other 60%."[24] EPA's CA measures need to be credible and useful to EPA managers and CA delivery agents so they can learn from their experience and make targeting decisions; they do not demand as high an accuracy rate as data needed for an enforcement case.

What follows is a brief list of some of the key feasibility, accuracy, and credibility issues identified for further consideration:

Diagnostic Capacity. EPA currently generates CA analytic reports manually. It needs computer-assisted analytic capacity to make it easier and less expensive to generate breakouts and look for patterns, correlations, and anomalies to tap the diagnostic value of measures.

Census of Assisted Facilities. EPA is currently able to measure the number of regulated facilities receiving direct assistance without a problem. It has far more difficulty measuring the number of regulated entities receiving assistance from the CA web-based Centers. EPA currently counts user sessions, but because search engines and other web-surfing tools generate a large and increasing number of hits to the site from non-users, user sessions are not a good indicator of the number of site users. Better methods need to be developed to assess how many regulated entities use the assistance centers, whether as one-time or repeat users. The number that sign up for a Center "list serve" might be a good indicator for some Centers, but not for Centers that include relevant environmental information in a general list serve to their membership.

---

[24] Kennedy School

Percentage of Regulated Facilities. To measure percentage changes in understanding, adoption of EMP, and reduction and proper handling of pollution among regulated parties, EPA would need to count or survey a statistically valid sample of all regulated parties. This would be prohibitively expensive to do for all sectors every year. This argues that, to get useful measures to assess program effectiveness and inform future action, it would be more feasible for EPA to focus its CA efforts on specific sectors, problems, or locations and measure changes within those areas.

Universe of Regulated Entities. Estimating the size of the regulated universe, especially for small business and the sector-specific CA centers, would improve the comprehensibility of EPA's current targets by better communicating how much of the problem EPA is tackling. Although estimating sector size can be difficult, EPA has made progress in that area in recent years.

Reporting Variations by Providers. There is significant variation in the number of regulated entities reached by each region, although less significant differences in outcomes. This may be due to differences in regional priorities, but may also result from variations in reporting practices. Reporting inconsistencies limit the credibility and the diagnostic value of data collected. If EPA revises its CA outcomes measures, more attention to reporting consistency would improve the data.

Reporting inconsistencies across similar programs can be overcome, however. If reporting is consistent within a region or other delivery unit across time, EPA can still use the measures diagnostically to learn by looking at the difference across regions. If one region (or center) shows a significant performance increase over time in one area that is not evident in other regions, it might reveal a positive deviant worthy of further attention even if they are measuring in different ways provided the internal reporting methods are consistent over time.

Under-counting and Over-counting. EPA's current measurement methods bring with them a number of inherent biases. The current method EPA uses to measure changes in outcomes associated with direct assistance is likely to under-report long-term effects. On-site observations and post-assistance surveys are more likely to detect immediate outcomes and less likely to detect outcomes that occur after the assistance is provided. Nor can the current measurement method detect whether or not reported changes are sustained. These are not all big problems, especially if the nature and size of the bias are consistent over time/ (If biases remain consistent, attention to the direction and magnitude of changes between time periods can be useful diagnostically.) EPA should be mindful of these biases, however, and minimize them to the extent possible.

EPA does not currently include a number of EPA efforts that are arguably compliance assistance, thereby undercounting the reach and effectiveness of the agency's CA efforts. Not included in the CA measures are the calls answered by the 800 numbers, much of the assistance provided by inspectors, assistance provided through sector-focused campaigns, and EPA-supported assistance work conducted by states and localities. This complicates the public's understanding of EPA's CA measures.

Response and Reporting Biases. Performance management experts warn against having those program delivery agents measure their own program results to avoid the temptation to manipulate the measures, especially if rewards or reputation are on the line.[25] Third-party observation also avoids the problem that people naturally tend to overestimate the value of what they have done.

Survey experts warn that survey respondents tend to answer questions in the way they think they are expected to answer them[26], while marketing experts find that people have a natural instinct to want to reciprocate positively to those who have given them assistance.[27]

Together, these tendencies suggest that CA survey respondents and CA provider-observers are likely to over-value of CA assistance generally, although asking questions appropriately can reduce this problem.

Information Collection Rule. To conduct a survey to collect information from regulated entities, EPA must get an ICR approval from OMB. This complicates EPA's CA survey efforts for longitudinal analyses, since the current ICR is time-limited. This requires EPA to submit a new ICR every few years to collect its CA measures. Data already collected by the agency, such as data on pollution and pollution incidents, and observer data do not require a new ICR so would be easier to use. Controlled studies or discrete surveys of awareness would also require an ICR. EPA and other agencies have received approval for these types of studies in the past, but it can be a long and difficult process.

F. Attribution

Many federal programs rely on others to accomplish their objectives. It is often hard to distinguish the effects of each contributing party to changes in outcomes of interest. The EPA compliance assistance program has two distinct attribution issues: one pertaining to other parts of EPA and one to states and localities. These attribution questions primarily interest budget offices. OMB has long pressed to understand whether and how much federal contribution is essential to program success and what proportion of each program's benefits should be attributable to the federal investment?

To deal with the CA attribution question within EPA, the agency has chosen to measure its compliance assistance programs separately from other compliance assurance programs. In principle, this allows EPA to isolate and measure the effect of its assistance programs separately from the effects of its monitoring and enforcement work. In fact, there are measurement challenges to be worked out, identified in the preceding discussion.

Now that EPA has started to measure compliance assistance across the country, regression analyses to assess the marginal contribution of assistance efforts to changes in

---

[25] Hatry

[26] Bradburn

[27] Cialdini

compliance and pollution outcomes may eventually be feasible, at least for facilities assisted that are in EPA databases. This would require identification of the assisted facilities, however, which is not currently being recorded in the EPA CA data base.

Replication efforts and controlled studies, even if not highly structured, are likely to be a better way for EPA to assess the marginal contribution of its assistance efforts compared to other compliance assuring activities.

EPA does not currently try to measure the effect of the CA it supports in the states, even though it is supporting numerous state CA efforts.

III. HOW AND WHEN COMPLIANCE ASSISTANCE DRIVES BEHAVIOR AND MOTIVATES CHANGE AND HOW THESE OUTCOMES CAN BE MEASURED
Cases referenced in this section are described in Appendix B.

This section briefly reviews lessons extractable from compliance assistance projects and compliance measurement projects around the country and world. To conduct this study, we reviewed measured compliance projects we were able to identify. We looked for information about factors that influence people to seek compliance assistance and circumstances that motivate them to change their behavior in an environmentally beneficially. We considered how compliance assistance drives behavior and motivates change. Finally, we looked for insight from these projects for measuring compliance assistance.

Our effort to review the state of practice clearly did not and could not capture all the work taking place around the U.S. and perhaps the world.

(In its initial questions to guide the project authors, EPA asked, "How can the influence of EPA assistance be separated/distinguished from the influence of state-provided assistance and other influences on regulated entities?" That question was addressed in the preceding section's "Attribution" discussion.)

A. Factors that Influence Regulated Parties to Seek Compliance Assistance?

Regulated parties are likely to seek compliance assistance if they (a) want to comply but do not know how to comply, (b) know that compliance assistance is available; and (c) believe it will help them comply. To answer the question, "why do people seek compliance assistance," we are primarily interested in the factors affecting their desire to comply, factors affecting their knowledge that compliance assistance is available, factors affecting their belief that compliance assistance will be helpful, and circumstances leading them to change their behavior. These factors and circumstances are summarized below.

- **Factors influencing regulated parties desire to comply:**
  - Economic utility, rational choice, deterrence theory [28]

---

[28] Deterrence theory is an economic model that assumes firms are rational actors and will comply with legal directives to the extent that the costs of expected penalties exceed the benefits of noncompliance (Kagan and Scholz 1984).

- o Social license utility: avoid harm to business reputation[29]
- o Personal health and risk utility
- o Strong sense of duty to obey the law[30]
- o Desire to be responsible environmental citizen[31]

Classic deterrence theory clearly explains some of the impetus to seek compliance assistance. Limited anecdotal and quantitative analysis suggest that an increase in inspection levels and enforcement seem to correlate with an increase in the number of regulated parties seeking compliance assistance. For example, very few participants signed up for one of EPA Region 1's sector-specific workshops until the region took an enforcement action in the same sector, after which the enrollment level rose.[32] In NH's generator study, reaching 10 percent of the regulated community with compliance surveyors conducting highly abbreviated facility reviews increased both the demand for and supply of compliance assistance. A study of Canadian wood industries found that, before inspections started, only ten to fifteen percent of the industry used best practices promoted for adoption. The rate of compliance increased to 80 to 90 percent after inspections began and warning letters started to be sent.[33] In other words, when regulated parties perceive a higher probability that their non-compliance will be detected, they seek more compliance assistance.

A few cases reviewed in Appendix B also identified a strong sense of duty and ethics as a reason some regulated parties sought CA.

- **Factors affecting their knowledge that compliance assistance is available:** A number of factors influence the chance that a regulated party will know that compliance assistance is available.
  - o **Turn on attention switch.** First, those offering compliance assistance must catch the attention of their target audience, i.e., they must "turn on the attention switch."[34] Factors that affect the attention switch include:
    - availability,
    - vividness, and
    - relevance of the compliance assistance information to the regulated party.[35]
  - o **Keep attention switch on/recall.** In addition, compliance assistance providers must maintain the attention of their target audience.[36] That is, the target audience must recall information previously received when it is relevant.[37]

---

[29] Gunningham et al 2004 329

[30] Tom R. Tyler, "Procedural Fairness and Compliance with the Law," *Swiss Journal of Economics and Statistics,* 1991, Vol 133 (2/2), 219-240.

[31] Hiscox, Michael J. and Smyth, Nicholas F.B. 2006. Is There Consumer Demand for Improved Labor Standards? Evidence from Field Experiments in Social Labeling. Working Paper. Department of Government. Harvard University. Cambridge MA 02138.

[32] E-mail exchange and conversation with EPA Region 1 staff.

[33] Krahn

[34] Attention switch is a term used by Wogalter and Vigilante in Wogalter 2006,

[35] J.B.T. Evans

[36] Attention maintenance is a term used by Wogalter and Vigilante in Wogalter 2006,

- **Factors influencing their belief that compliance assistance will be helpful**: People in organizations are more likely to believe a new practice will be helpful to them when they see it has been helpful to others.[37] Potential adopters' belief that something will be helpful is likely to be stronger when they are part of an existing interpersonal network with others who have adopted the new practice (early adopters), especially when early adopters are seen as opinion leaders. Also, beliefs and practices transfer more quickly among those who are similar and geographically proximate.[39]

Cooperation with industry associations or regulated parties themselves seems to be helpful in getting people to seek compliance assistance. Several projects found that trade association partnerships helped surmount two-way communication barriers, where the agency does not understand regulated entities and visa versa. While w found no quantitative assessment of the marginal impact of involving a trade association, several projects identified it as a useful way to reduce distrust, a potential barrier to CA receptivity.

B. How and When Compliance Assistance Drives Behavior and Motivates Change

To understand how and when compliance assistance drives behavior or motivates change and how these outcomes can be measured, we reviewed a large body of the compliance literature. Three cases found a likely positive impact of compliance assistance on compliance. Only one, however, delivered compliance assistance without also conducting modified inspections. Another study with a very small sample tried to compare changes in compliance with and without assistance. It is most useful for suggesting the possibility of conducting a randomized assignment experiment. Other studies noted changes in intermediate outcomes, including an increased awareness of regulatory obligations, measured by increases in the number of permitted parties after agency notification of specific regulatory obligations. In addition, the lack of understanding of regulatory obligations appears to be a significant barrier to compliance.

The state of Washington carried out a project that was designed in part "to compare technical assistance visits with formal inspections to see how these two types of facility visits compared in obtaining and maintaining compliance over a similar time period." The study found that "[f]or facilities generating relatively small amounts of hazardous waste, compliance technical assistance visits …appear to be as effective as formal CEI inspections in maintaining regulatory compliance." For relatively small generators, the report recommended continuing "technical assistance" visits, concluding they can achieve results equivalent to formal inspections, with fewer resources.

There is also evidence that compliance surveys (minimal inspections that look only at a limited number of facility practices and conditions) can effectively motivate improved

---

[37] Wogalter and Vigilante in Wogalter 2006.
[38] Kelman 2005, p.128.
[39] Rogers

compliance, especially when paired with CA handouts. New Hampshire used this approach to deal with its chronic shortage of RCRA inspectors. It trained and hired summer interns for compliance surveys over a six week period. The "compliance surveyors" visited 439 facilities in 10 weeks, compared to the 306 facilities NH had inspected in the prior ten years. The enforcement threats were clearly low in the NH RCRA program. After the surveyor visited, NH saw a large rise in requests for the CA information handed out to visited facilities, suggesting a positive social networking effect was achieved by reaching approximately 10% of the regulated universe. Following the compliance surveys, a trade association also hired a full-time specialist to conduct semi-annual on-site compliance evaluations for its members and a supplier started sending letters to clients on proper management and disposal practices. In addition, compliance surveyors contacted regular inspectors when they saw serious problems. In other words, they served not just as surveyors but as compliance scouts.

Following NH's lead, the state of Connecticut worked with college interns to conduct modified, simplified inspections (compliance indicator surveys) at small quality generators of hazardous waste. "During the summers of 2004 and 2005, Department staff [including interns] conducted compliance indicator surveys at a total of 1,173 SQGs," which represented a larger percentage of the regulated universe than state inspectors could address. "The site surveys for SQGs consisted of 10 questions designed to assess limited areas of compliance considered indicators of overall compliance. The average overall compliance rate for all 10 survey questions was 75% in 2004 and 81% for 2005. The benefits realized from implementation of the initiative include an increased field presence, the ability to screen sites for full inspections, the development of compliance rates and measures, identification of areas where additional compliance assistance is needed as well as identification of areas where enforcement action is necessary.[40]

In another case, compliance assistance paired with pre-notification of inspections did not cause a marginally greater increase in compliance than pre-notification of inspections alone. In this project, 7 of 11 facilities receiving compliance assistance materials said they had read them, but overall facilities receiving compliance assistance did not improve their compliance rates more than those that did not. This suggests that: a) facilities may overstate the degree to which they read compliance assistance materials, b) facilities read but did not absorb compliance assistance information, c) facilities read and absorbed compliance assistance information, but were not able to translate this information into improved compliance, and/or d) facilities are motivated enough to comply, they will find a way to do so, even without compliance assistance. It may also be true that the way compliance assistance was provided in this case (written materials) was not effective; perhaps the materials were not well designed, or perhaps direct compliance assistance would have been more effective. Note that the sample size for this project was small, so it is not advisable to weight these conclusions too heavily.[41] This study is primarily useful for suggesting both the possibility of conducting a randomized assignment experiment and the potential value of more focused inquiries on the reasons for non-compliance.

---

[40] CT STAG grant report, p. 5-6.
[41] CO Compass project – SQG sector

As suggested in the logic model above, awareness of compliance obligations is likely to be an essential early step to getting regulated entities to change their behavior and reminders of regulatory obligations can be very useful, especially for general permit holders and other small business. A number of projects found high numbers of "non-filers," facilities that should have been regulated and were not among general permit holders and other regulated small businesses. Notifying non-filers via mail seems an effective means to increase the number of registrants/permit holders,[42] even without perceived imminent threat of inspections or enforcement.

Understanding compliance obligations is also necessary for improved compliance. A number of projects found it necessary to revise their initial CA material so regulated parties could understand it. In its auto body and repair project, the state worked with the target regulated community's to revise the material so it would be understandable. The state needed to schedule time to revise it, and worked closely with the regulated entities to assure they understood the material.

Even with awareness and understanding, people and regulated parties do not always change their behavior. Physicians have known for over a hundred years that proper hand-washing saves lives, yet tens of thousands of Americans every year die of infections caused by improper hand-washing methods.[43] Motivating behavior change among this highly educated population, even when they are aware of a problem, has proved daunting. What prompts people to comply? Measuring awareness and understanding may prove a useful intermediate outcome measure, but in most cases, may need to be paired with end outcome measures to assure progress toward compliance and reduce pollution impact does not stop at the intermediate stages.

C. Motivating Compliance and Behavior Change

Although the environmental and regulatory literature about using compliance assistance to motivate behavior change is limited, the social psychology literature is rich with studies and insights about the "psychology of compliance" that can help answer the question: "What are the factors that cause one person to say yes to another person?" without first making a utilitarian calculation. They have identified six basic categories, psychological principle that directs human behavior, that prove to be effective compliance-inducing tactics: reciprocation, consistency, social proof, liking, authority and scarcity, which are described in turn below:

- **Reciprocation.** People all over the world have an instinct to repay what has been given them. Free address labels employ the reciprocity principal to boost donations. One non-profit organization saw its response rate double after including individualized address labels in its mailings.[44]

- **Consistency:** Once someone commits to a position, they tell themselves they made the right decision and want to be consistent with it. This principle has made

---

[42] CT General Permits
[43] Ayres
[44] Cialdini

the "foot-in-the-door" sales technique very popular; once a small commitment is made, it is easier to get a larger commitment.[45] This principle can be applied to public policy problems. For example, researchers knocked on doors of randomly selected houses in a neighborhood and asked people to sign a petition or place a sticker in their car window supporting safe driving. Two weeks later, researchers visited the same houses and a second set of randomly selected houses not previously visited and asked them to put a poorly produced sign in their yard, "Drive Carefully." Seventy percent of those who had previously agreed to sign the petition or place a sticker in their car window agreed to put up the yard sign, but only seventeen percent of those not previously visited agreed.[46]

An Iowa research project used the consistency principle to increase energy conservation. In the first round of an experiment, researchers knocked on doors and obtained commitments to conserve energy and provided one-on-one direct energy-saving assistance. The results were that energy consumption levels, as measured by utility bills, did not change. In the second round of the experiment, researchers used the consistency principle to increase energy conservation. This time, the assistance provider promised to list participant names in the paper as fuel-conserving citizens. Within a month, energy consumption levels dropped 12.2 percent. Even after participants in the second round of the experiment received a letter informing them that their names would no longer be published in the paper, they continued to save energy, increasing the conservation rate to 15.5 percent for the duration of the winter. The perception of themselves as energy conservers, reinforced by their knowledge that others also saw them as conservers, provoked significant behavior change. EPA compliance assistance efforts could similarly experiment with ways to build the commitment.

- **Social Proof:** The inclination to follow the leader, previously discussed, can motivate compliance. Children with a fear of dogs can lose those fears simply by watching other children play happily with dogs, even if viewing only a film clip. Film clips showing a variety of children happily playing with dogs are more effective than those showing just a single child.[47]

  With social proof, similarity matters. People are more likely to follow others whom they see as similar to themselves. Ambiguity also matters. The more ambiguous the situation, the more the social proof principle will apply.[48]

  Social proof often operates among organizations, but not always as predicted. Organizations tend to imitate other organizations they deem as exceptional, even when they are exceptionally bad, but not those they consider average. They also

---

[45] Cialdini
[46] Freedman and Fraser
[47] Cialdini
[48] Cialdini

tend to imitate larger firms and follow a specific strategy widely used by other firms in the same industry.[49]

The social proof principle suggests one compliance assistance tactic worth testing: identify and publicize (or hire models) with whom the target audience identifies to model compliant behavior.[50]

- **Liking.** As noted earlier, we are more likely to say yes to people whom we like. A number of factors influence our level of liking, including looks, similarity, compliments (provided they feel genuine), cooperation that achieves shared success, and association. The liking principle suggests that attention should be paid to the way compliance assistance delivery agents dress, how they make compliance assistance recipients feel, and associating compliance assistance with something the target audience already likes.[51]

- **Authority.** Milgrom's famous experiments showed that people will defer to the command of a higher authority, even when asked to do something against their own best judgment. Compliance assistance providers naturally tap into that deference to authority.

- **Scarcity.** People want things more when they think they are in short supply. This has multiple implications for compliance assistance providers. First, it suggests that those marketing compliance assistance may want to try marketing it as a scarce good, suggesting, for example, that "a top compliance assistance expert from Washington will only be available in a specific area for two days in March." Second, it suggests that compliance assistance providers should think carefully about, and test, different ways to frame information about compliance assistance, compliance obligations, and the threat of non-compliance to give it the highest perceived value and motivate compliance.

D. Is there any Significant Variation in Changes in Environmental Management Practices and Pollution Reduction Outcomes based on the Type of Compliance Activity,

Unfortunately, the case studies we found do not directly compare the effectiveness of one type of compliance assistance against another. However, the literature does suggest that compliance assistance can be effective.

There are tradeoffs between providing indirect vs. direct assistance. For example, indirect assistance may seem to take less time and impose fewer opportunity costs on small businesses, and therefore may have greater participation. Indirect assistance is also valuable because it can be offered anonymously. However it is not clear whether indirect

---

[49] P.R. Haunschild and A.S. Miner, "Modes of Interorganizational Imitation: The Effects of Outcome Salience and Uncertainty," *Administrative Scien Quarterly*, 42, pp.472-500, cited in Argote.
[50] Cialdini, Ch. 4.
[51] Cialdini, Ch. 5

assistance is as effective as direct assistance in achieving compliance. For example, in the case of the Massachusetts Pesticide STAG project, indirect assistance was effective in raising awareness – "Using a questionnaire, the audience was asked to answer four questions (two legal & two practical) before the presentation and then an additional four questions after the workshop. On average, the percentage of questions answered correctly before the workshop was 78% and after the workshop the percentage of correct answers increased 11 points to 89%." However, direct assistance was more effective in actually achieving compliance: "Of the five facilities which were completely out of compliance, six were schools and seven were daycares. There was no correlation between attendance at the workshops and compliance – five of the schools had attended the trainings but were not in full compliance. After the follow-up visits by field staff, however, full compliance was attained by all the targeted facilities. The value of intervention by field staff is thus highly significant. A direct correlation between the understanding of IPM and participation in the workshops was not possible to establish, however it did appear that attendance at the workshops led to a higher score on the quiz." (MA Pest STAG grant)

Colorado SCORE project found that mailings were not as effective as on-site assistance and consultation, although CT General Permit project found that mailing could be effective, especially when paired with enforcement. In one project that used mailings with enforcement, compliance levels went up. In a second project designed to bring non-filers, entities that were regulated and should have been registered and reporting, mailings more than tripled the number of regulated parties.

E. How Compliance Assistance Outcomes can be measured

Several experimental efforts tried to develop better ways to measure the effect of compliance assistance and then to assess before-and-after affects of different approaches involving CA. Almost all of them paired compliance assistance with other compliance assurance tools. Key insights from those projects are summarized in the narrative below. The studies are presented in greater detail in Appendix B.

To assess the effectiveness of compliance assistance, it is necessary to be able to measure changes in outcomes. Significant progress has been made in recent years developing methods for measuring compliance assistance and behavior change. The methods being developed to measure behavior and compliance (and changes in them) vary along a number of dimensions: what is counted, who is inspected, and how they are inspected.

Scope of Measurement. A few states have developed the capacity to calculate compliance rate for all programs. They do not use census or probability sampling, but report compliance rates for facilities monitored together with compliance monitoring rates. These states produce annual reports to the public showing trends in both areas. In programs with high inspection rates, these systems could be used to test the effectiveness of different compliance assistance approaches.[52]

---

[52] NJ, CT, and MD. See Metzenbaum, et al, "Memo on Measurement" at www.complianceconsortium.org

Most environmental agencies that have tried to measure compliance program effects target a specific sector or program. Some then measure all facilities in a sector or program[53]; while others conduct a probability sample.[54]

What is Measured? Many agencies have tried to measure compliance. A few have tried to measure behavioral change. And a very small number, via survey and not observation, have measured attitudes and awareness.

Different metrics are used to measure compliance. Many agencies count percentage of facilities exceeding some, sometimes risk-adjusted, threshold number of non-compliant incidents (often one) as non-compliant. For agencies administering compliance assistance, this proved too rough a gauge to reveal information about the degree of non-compliance and, after assistance, the degree of improvement. To address this problem, a number of agencies with high monitoring levels have developed systems to count the frequency of non-compliance (counting unwanted events) and found this useful both to assess trends and to dig down diagnostically.[55]

A number of projects, including the NH, CT, and WA programs described above, identify subsets of compliance obligations to be counted rather than the full set. NH measured not only compliance indicators and but non-mandated beneficial environmental behaviors.[56]

A few environmental agencies have started to measure intermediate outcome indicators, including awareness, attitudes, and understanding. Oregon and King County, WA have both used random surveys of the population to measure awareness and attitudes of regulatory obligations and behavioral practices. King County, WA has surveyed for more than one year, so it is beginning to see trends. It also surveys to ask residents about behavioral change in their yard management practices. Oregon conducted two surveys, one of the general population and one of regulated parties. The UK Environment Agency has conducted surveys that found that ¾ of SMEs think that they do not have a negative impact on the environment and are not well-versed in environmental legislation.

Connecticut also conducted a project that measured changes in understanding in the auto recycling industry. The state distributed guidance materials to all known auto recyclers in the state in 2004, and conducted four training sessions that covered sections of the guidance manual in detail. The state developed a questionnaire to assess understanding of requirements and compliance before and after this compliance assistance intervention, in order to measure its effect. "Prior to [distribution of the guidance materials and] the training (2003), auto recyclers did not have an understanding of what their operating status was related to their hazardous waste generator status. They did not know if they were operating as a large quantity, small quantity or conditionally exempt generator. Following the [distribution of the guidance materials and] training (2005), operators had a better understanding of operating status. In 2003, 19% of auto recyclers responding

---

[53] CO RCRA traditional.
[54] CA ARB, CO COMPASS for asphalt plants, SQG.
[55] Metzenbaum Spring 2007, Shewmake
[56] NH

identified their facility as operating as a CESQG and 43% did not know their status. In 2005, following the training, 80% identified their facility as operating as a CESQG and only 10% did not know their status.

Could EPA's current pollution reporting requirements be used as a CA measure? Wisconsin recently attempted to compare pollution trends of participants in the Wisconsin Green Tier program, its performance-focused regulatory program with pollution trends of other facilities in the state.[57] Although Green Tier is not a compliance assistance program, this informal analysis suggests a possible outcome indicator useful for diagnostic analyses to assess the effectiveness of CA activities on waste water discharges and emissions, should EPA decide to direct CA efforts toward DMR and air emission reporters.

Similar to the pollution incidents reports used by the Coast Guard and several other federal agencies, Michigan DEQ produces a Combined Sewer Overflow (CSO) and Sanitary Sewer Overflow (SSO) Annual Report. The 2005 report shows the number of SSO/ CSO events per year and SSO/CSO total volume per year. The report shows the extent to which sewage was released into the environment and progress reducing overflow events. It also indicates the number and volume of flows associated with specific facilities, and identifies facilities with the greatest negative impact.[58] Other agencies also report pollution incidents in various ways. Illinois alerts citizens to sites and facilities with off-site soil or groundwater contamination that could affect public health.[59].

New Mexico counted number of facilities added to the Class V UIC inventory and the number of new permits issued as two of several outcome indicators to gauge the effectiveness of its compliance assurance work.

Units of Measure.
For its hazardous waste program Washington State experimented with three different, compliance indexes ("score") for regulated generators: Y/N, Categorical, Absolute.

Who does the inspection? Some states used their regular inspectors to conduct inspections, some tried self-reporting, one hired a third party, and a few have turned to trained interns to conduct "compliance surveys."

States using their regular inspectors to measure compliance have, on occasion, encountered problems.[60] Their traditional inspectors were easily distracted by problems

---

[57] http://dnr.wi.gov/org/caer/cea/ecpp/reports/2006annualreport.pdf

[58] http://www.deq.state.mi.us/documents/deq-wb-csossoreport05.pdf

[59] http://www.ecos.org/files/2781_file_Spring_2007_ECOStates.pdf?PHPSESSID=6133e12b8cbad7d0dbff0826c348b1e8

[60] MD

they wanted to follow up on. They got frustrated by restrictions on them just to conduct a monitoring inspection.[61]

Several states, including the CO SCORE program and numerous ERP projects around the country, are testing the use of self-reporting checklists, supported with a CA workbook. Regulated parties tally their own compliance and behavior change and report it to the government. Massachusetts, which piloted this approach and requires the use of the workbook in its Environmental Results Program (ERP), mandates that the workbook be filled out as a regulatory obligation and reporters certify to its accuracy. Maryland, in its auto body and repair project, tried something similar and concluded its program would have been far more successful if it had made workbook use a regulatory obligation, not just a compliance assistance tool. Some of the regulated parties in the CO SCORE project offered comments that the checklist was a helpful refresher on hazardous waste regulations.

Connecticut used a consultant to help it measure compliance and the nature of non-compliance among general permit holders.
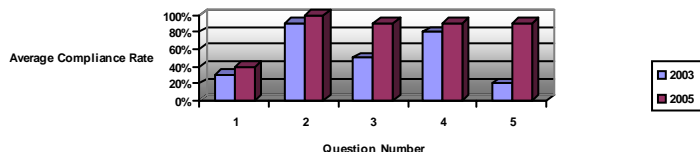
The NH approach described above appears to be one of the most promising compliance monitoring practices, in part because it is so economic both in its use of observers and in the time needed for observation. NH's program adapted the MA ERP measurement approach to a traditional program that did not use self-reporting. It drastically cut the number of items on the inspection check list and, as noted above, used trained interns to conduct simplified compliance monitoring survey rather than full inspections. Colorado used a similar approach to follow-up on the effectiveness of its CA work with asphalt plants.[62] It concluded this monitoring format was easy for reviewers to understand, even if they had never conducted inspections and were not familiar with the inspection process. In other words, this approach may hold promise as a compliance monitoring method. Consider the benefits of this approach: a possibly greater deterrence effect in programs with low inspection rates, increased demand for CA, relatively low costs, and possibly higher detection rates for serious non-compliance.

Uses of Measures. As noted in section III.B, a small number of state agencies have begun to measure inspection rates and compliance rates, in a non-statistically valid way, to report consistently to the public on an annual basis. In addition, several agencies have begun to use their measures not just to gauge overall program assistance, but for diagnostic and learning purposes. CO used its data to find an anomaly, low compliance levels on an easy compliance obligation, which it could easily fix. NH conducted a factor analysis of its results to search for correlations between inspection violations, building on earlier work done by Colorado. WA State conducted breakout analyses on its data to discover, to its surprise, that LQGs had worse baseline compliance levels than MQGs. It also found, to its relief, that the lowest type of non-compliance for LQGs and MQGs was spills. CT analyzed data it collected on behavior change diagnostically to understand where it was making progress and where problems remained. It asked five

---

[61] MD
[62] CO COMET

questions about specific behavioral practices for a specific sector, such as "Is crushing performed on an impermeable surface?" and "Are batteries stored on an impermeable surface?" and then constructed a bar chart to show where problems were the worst and progress the greatest.

Summary

Recent efforts to measure compliance and other behavior changes have generated many promising insights and more sophisticated measurement methods. The studies are beginning to reveal useful insights about changes in one place over time. In addition, they are paving a path that may help EPA develop a more robust capacity to measure nationally and learn from sub-national experiments to improve the quality of EPA and others' CA programs. Given that much of the recent compliance measurement projects, received financial supported from EPA via competitive grants, this seems a promising strategy to consider for further support in the future.

IV. RECOMMENDATIONS

EPA has made great progress over the past decade in the development and collection of better measures to advance Goal 5 of its Strategic Plan, including its compliance assistance work. This white paper seeks to provide EPA with additional background information to help it understand the state of science and practice in CA performance measurement. Based on this review, we offer the following recommendations to EPA.

Comprehensibility and Coherence

We reviewed EPA's current GPRA CA measures in the context of five key criteria. We found them conceptually comprehensible and coherent. They align with the agency's mission and strategic goals. They also reflect an understanding of behavioral science's findings about the likely intermediate outcomes needed to reduce pollution risks. Adjustments are advised, however, to address three problems with EPA's current CA measures: a confusing calculation of program impact, insufficient focus to communicate organizational priorities, and the need for context to understand the import of the targets.

Because the cost of measuring the whole regulated universe and changes within that universe are prohibitive, EPA sensibly opted not to attempt the costly count of changes among regulated parties. Instead, it opted to measure changes among those assisted. This measure confuses understanding of the impact of EPA's CA work. If the nature and number of those receiving assistance increase significantly but with a slightly lower impact per person, EPA could successfully raise understanding and behavior among more people, but earn a lower CA score.

> **Recommendation 1: We recommend that EPA consider an adjustment to its current CA measures to address this problem. One way it could try to do that would be multiplying the percentage change in each category by the number assisted in that category wherever known to calculate a baseline-adjusted indicator of improved outcomes among regulated entities. This will make trend comparisons more accurate.**

To adjust its CA national measure as suggested in Recommendation 1, EPA will need to have a better estimate of the number of Center users, not the number of user sessions.

> **Recommendation 1a: We recommend that EPA consider working with the Centers, its web-consultants, OEI, and other experts to develop better ways to estimate the actual number and composition of web-based assistance users.**

When measures are too general, they are not effective at focusing agency action. Nor can they clearly communicate priorities to the public about agency priorities. GPRA measures used by other federal agencies tend to be more specific, helping agency staff better understand where to focus attention.

> **Recommendation 2: We urge EPA to consider adopting more specific CA targets that clarify the priority types of understanding, behavior change, and pollution reductions to be improved and the target audiences, whether by sector, regulatory program, or geographic location.**

The comprehensibility of EPA's CA annual measure would be enhanced with information about the size of the universe EPA is trying to reach, not as a CA measure but to provide a context for understanding the import of the measures EPA is reporting and to inform resource allocation decisions. EPA already produces some estimates of sector size.

> **Recommendation 3: We urge EPA to consider producing estimates of the size of the regulated target audience it is trying to reach to better inform the public and its own decision-makers about appropriate CA targets and the likely impact of its strategies relative to the size of the problem. If EPA specifies its targets as suggested in Recommendation 2, we urge it to provide estimates of the size of the universe for each target population it has chosen.**

Motivation

Based on the strong findings in the cognitive psychology literature about the motivational power of specific, challenging targets, we would suggest that that EPA's current national CA targets could be more specific and challenging.

We found that many other federal agencies adopted more specific targets, and believe this would not only enhance the motivational value but also the comprehensibility of EPA's current targets. In addition, we found examples of federal agencies using national sectoral measures as an effective motivator of regulated parties.

> **Recommendation 4: As stated in Recommendation 2, we suggest EPA consider making its measures more specific by setting specific CA targets by sector, program, region, and Center. This will also have motivational value.**
>
> **We would also urge EPA to consider adopting a more challenging, albeit realistic target. It can do that by setting performance targets higher than current levels once it is assesses its current performance levels more accurately and possibly for selected sectors or programs. It might also experiment with setting more challenging targets by identifying and setting a target to raise performance levels in low-performing areas or by adjusting EPA's current measures of CA changes from a binary Y/N value to a multiple-value scale that reflects more finely the degrees of change.**

Performance measures have a stronger motivational effect when those EPA is seeking to motivate find the measures relevant and useful to their own work, not just as a reporting requirement.

> **Recommendation 5: If it is not already doing so, we would urge EPA to set discrete CA targets, by sector or program, and measure progress for each major delivery unit which then get rolled up to national measures. Also, to figure out if CA delivery agents find the CA measures useful, we would urge EPA to engage the regions and other CA delivery agents in a broader discussion about whether or not they find the CA measures useful.**

Diagnostic Value

Measures are useful diagnostically to illuminate what works and what does not. Measures can also illuminate problems that need attention and their relative import. Using measures this way supports learning and continuous improvement. National measures, per se, have limited diagnostic value. Their real motivation and diagnostic value arise when national measurement is disaggregated to reveal performance variations across the segments.

> **Recommendation 6: EPA should consider, as soon as feasible, starting to experiment with break-out analysis of EPA's current CA database to see if it can find interesting patterns, variations, and anomalies that help it identify problems and leading it to the discovery of promising practices. EPA is encouraged to search for "positive deviants," those who outperform others (not just because of reporting differences.)**

EPA currently generates all of its CA analytic reports manually. It lacks computer-assisted analytic capacity that would make it easier and less expensive to generate breakouts and look for correlations and anomalies.

> **Recommendation 6a: We would urge EPA to examine ways to enhance its capacity to generate more diagnostic analyses of the data it collects on a regular basis.**

EPA has supported discrete diagnostic studies and experiments in the past to complement its national CA measures. These studies have proven extremely beneficial in advancing measurement methodologies and advancing understanding of the effectiveness of assistance and other compliance assuring efforts.

> **Recommendation 7: EPA should consider funding more studies, including randomized assignment studies (with randomized tests of different web-based strategies), to refine understanding of how variations in types of compliance assistance affect outcomes. It should consider experimenting with the approach used by NHTSA of using competitive grants to enlist other governments to participate in random assignment trials. This could be especially beneficial in helping EPA, states, localities, and tribes to identify not only more effective, but also more cost-effective compliance-assurance approaches.**

Common measures across states would be helpful for inter-state diagnostic analysis. EPA is currently supporting a state-led initiative in the Northeast involving 10 states to develop common performance indicators for specific sectors, starting with small quantity generators and auto body shops. This effort builds on the Massachusetts ERP measurement method, as well as the measurement efforts in New Hampshire and Connecticut supported by EPA.

> **Recommendation 8: It is hoped that EPA will continue to support the work of the Northeastern states to standardize compliance measurement for two sectors, and that it will support expansion of this work to other states and to other sectors if it is successful. It is also hoped that EPA and the states will align this work with Exchange Network efforts.**

EPA's CA planning and resource allocation decisions would be greatly enhanced by better understanding, on an ongoing basis until high awareness rates are firmly established, of the awareness of the availability of CA among its target audience.

> **Recommendation 9: EPA should consider conducting ongoing studies to gauge awareness of its CA activities among target regulated parties. It could consider using its previous study of awareness among the regulated community of two CA centers as a baseline to determine if awareness has increased for those Centers, as an initial indicator of change.**

Replication Value. The review of the compliance assistance literature, especially reports on regional and STAG grant-funded compliance assistance and measurement projects, suggests that a tremendous amount of beneficial CA work is taking place across the country. Several before-and-after studies of compliance assistance efforts, report noteworthy improvements in compliance levels, EMP, or pollution. It also suggests that EPA's investment in state measurement experiments had a high payoff in knowledge and the advance of measurement methods and practices in the states. But many questions remain to be answered. One key question that is still hard to answer is: what types of CA are worth replicating because they are comparatively more effective and cost-effective than other approaches?

> **Recommendation 10: EPA is urged to consider adopting a more systematic strategy to search for successes, in the regions, states, localities, and tribes, and promote their replication. This would include a more systematic approach to tracking and sharing information about statistically valid before-and-after or random assignment studies. (The authors of this report developed a format for organizing the project-based studies we reviewed. We would be glad to provide the agency the template we used if EPA chooses to pursue this recommendation.)**

> **Recommendation 11: In addition, EPA is urged to consider repeat funding of more compliance and compliance assistance measurement work.**

Feasibility

EPA has a number of feasibility issues with which to contend. We will not offer additional specific recommendations about minimizing statistical problems, so offer here only two other recommendations, one pertaining to ICRs and one to the use and usefulness of the measures to CA delivery agents.

> **Recommendation 12: We would suggest that EPA consider thinking, over the longer term, about its ICR needs in terms of how it wants to use the data it collects diagnostically to better understand specific and important gaps in awareness and gaps in understanding, and to aid the search for positive deviants.**

Better Measures?

With the exception of the IRS which periodically measures compliance in a statistically valid way but does not use compliance as an annual measure, the other federal regulatory agencies reviewed for this report that deliver compliance assistance integrated the measurement of their compliance assistance efforts with measurement of their other outcome-improving activities. They used one national measure to capture the work of multiple program functions, rather than trying to gauge and attribute the distinct contributions of different offices with a shared objective. Interestingly, a large number of agencies use a similar outcome indicator applied to each of their own areas, the number of unwanted incidents.

This raises a question about the feasibility of EPA similarly adopting shared outcome measures to inform and guide its work across program functions, or at least initially across compliance-assuring functions. Also, given the success other agencies have had measuring unwanted incidents as a powerful and useful measure that helps them find problem, set priorities based on relative risk, and find promising practices that drive the number of unwanted incidents down when replicated, it seems an avenue worthy of further exploration. EPA and states already report pollution incidents in a number of ways, including the CAA Section 112(r) data, SSO/CSO overflow events, and permit exceedance notification systems. The question is, can that data be used and enhanced in ways that will make it more useful?

> **Recommendation 14: EPA should consider developing a pilot program to test the feasibility of counting the number of pollution incidents as a CA and perhaps an integrated program performance measure. Assembling this information would require significant effort on the part of EPA at the national level, but we suggest EPA consider exploring this option further to open up the possibility for adoption of this measure in the future. Given the success of its prior grants in stimulating progress among the states in compliance measurement and compliance-boosting experiments, EPA might want to consider using competitive grants to states or even localities to see if any might want to try an experiment to develop and test a system to use pollution incidents as a program performance measure. EPA might also consider soliciting interest from one or more regions on this effort. If it pursues this option, it would most likely be more attractive to a region if it could make pollution incidents a primary performance measure in the EPA Annual Commitment System, an "instead of" measure and not an "in addition to" measure. If EPA opts to pursue the recommendation to set region and sector specific targets, it might explore the feasibility of testing the use of this measure in selected regions for selected sectors.**

Our review of state and local practices identified another potentially promising measure EPA might eventually be able to use as a national measure for selected targets. This is the use of trained but relatively inexpensive personnel to conduct "compliance surveys" focusing on a relatively small (15? – 10 common across states and time and 5 experimental or state-specific) number of compliance and business practice indicators at randomly selected facilities. This approach could be an affordable and effective way to

measure compliance and business practice changes annually to note compliance trends and identify more prevalent compliance problems among selected programs and sectors. This is the approach NH used with its LQGs and SQGs. This approach not only produced a statistically valid compliance rate, it also increased awareness of the agency among regulated entities in a program with a low inspection rate (SQGs), stimulated demand for increased compliance assistance information, motivated the trade association and a supplier to provide more assistance, and provided a "scouting" function to alert inspectors when serious problems were seen at facilities visited.

> **Recommendation 15: EPA should consider supporting a project to enlist regions and states in other regions to test replication of the "compliance survey" approach and, if successful, roll it out for national adoption to develop state and national compliance rates for a single program or sector. CT has already demonstrated the replicability of the approach NH developed. EPA should also examine if and how the NEWMOA Common Measures project might support this.**

Finally, over the last few decades, EPA has adopted an increasing number of projects, especially in the air and water areas, focused on reducing pollution loads and improving ambient conditions in specific places. These projects often integrate a large variety of agency activities, with the mix changing over time, all coordinated by a focus on improving measures of ambient conditions in a specific location. It is challenging to think about a national CA measure with its own ambient condition target, but not inconceivable to think about CA playing a leadership role managing across functions to achieve an ambient target for a specific area.

> **Recommendation 16: EPA should consider adopting, among its CA measures, site specific ambient targets for places where significant improvements in compliance and better environmental practices could bring about improved environmental conditions.**

A shift to CA measures more focused on increasing compliance among all regulated parties, reducing pollution events, or improving environmental conditions would not be easy or quick. The Coast Guard reported nearly fifteen years ago about the difficulty of its multi-year effort to shift to outcome-focused measurement, but it also reported at the time how much the shift stimulated innovation and discovery of more effective and flexible approaches that seamlessly integrated enforcement and assistance tools.[63] The success of the Coast Guard approach is evident today in its marine environmental protection and vessel safety work. EPA has made so much progress over the last decade testing projects and measures more focused on outcomes that it appears it may be ready to take this next big step.

---

[63] "Stimulate discovery" was not used as a criteria for this review of EPA's CA measures, so the discovery-stimulating value of outcome-focused measurement was not discussed in the text of this paper. The value of outcome focused goals has been examined by many, including March and Kowalewski, and underlies the adoption of GPRA.

V. BIBLIOGRAPHY

The American Customer Satisfaction Index. (2007). *ACSI E-Government Satisfaction Index*. Retrieved from http://www.theacsi.org/images/stories/images/govsatscores/e-gov_Q3_2007.xls.

Andrews, R.N.L. (2005). *Environmental Management Systems: Do Formalized Management Systems Produce Superior Performance? – Final Report*. Paper prepared for the U.S. Environmental Protection Agency, EPA Grant ID # R829440.

Argote, L. (1999). *Organizational Learning: Creating, Retaining, and Transferring Knowledge*. New York: Springer.

Ayres, I. (2007). *Super Crunchers: Why Thinking-by-Numbers is the New Way to Be Smart*. New York: Bantam.

Ayres, I., & Braithwaite, J. (1992). *Responsive Regulation: Transcending the Deregulation Debate*. New York: Oxford.

Baldwin, T.T., & Ford, K.J. (1988). Transfer of Training: A Review and Directions for Further Research. *Personnel Psychology*, *41*, 63 – 105.

Bardach, E., & Kagan, R.A. (2003). *Going by the Book: The Problem of Regulatory Unreasonableness*. New Brunswick: Transaction Publishers.

Belfiore, B. [Benchmark Portal], personal communication, September 19, 2007.

Bennear, L.D.S. (2003). *Are Management-Based Regulations Effective?: Evidence from State Pollution Prevention Programs*. Regulatory Policy Program Working Paper RPP-2003-21. Cambridge, MA: Center for Business and Government, John F. Kennedy School of Government, Harvard University.

Bosnjak, M., & Tuten, T.L. (2001). Classifying Response Behaviors in Web-based Surveys. *Journal of Computer-Mediated Communication*, *6,* (3). Retrieved November 18, 2007, from http://jcmc.indiana.edu/vol6/issue3/boznjak.html.

Bradburn, N. et al (2004). *Asking Questions: The Definitive Guide to Questionnaire Design – for Market Research, Political Polls, and Social and Health Questionnaires*. San Francisco: Jossey-Bass.

Braithwaite, V. (2002). *Taxing Democracy*. Hants, England: Ashgate Publishing Ltd.

Braithwaite, V., & Braithwaite, J. *An Evolving Model for Tax Enforcement*.

Buckhurst, K. [Accenture], personal communication, October 1, 2007.

California Environmental Protection Agency, Air Resources Board. (2003). *USEPA/OECA/OC State and Tribal Assistance Grants (1999-2000), California: Air Toxics Compliance Monitoring Report.* Retrieved September 13, 2007, from http://www.epa.gov/compliance/resources/reports/stag/fy2000/stag00-ca-airtoxics-final.pdf. Paper prepared for the U.S. Environmental Protection Agency, EPA Grant ID # 98951301-0.

Carlough, L. *General Deterrence of Environmental Violation.* Portland, Oregon: Oregon Department of Environmental Quality.

Cialdini, R.B. (2001). *Influence: Science and Practice, Fourth Edition.* Boston: Allyn and Bacon.

Coglianese, C., & Kagan, R.A. (2007). Introduction. In C. Coglianese & R.A. Kagan (Eds.), *Regulation and Regulatory Processes* (pp. xi-xxvi). Burlington, VT: Ashgate Publishing Company.

Connecticut Department of Environmental Protection. (2004). *USEPA/OECA/OC State and Tribal Assistance Grants (2000-2001), Connecticut: Developing Compliance Measurement and Management Strategy for Connecticut's General Permit Program – Final Report.* Paper prepared for the U.S. Environmental Protection Agency.

Connecticut Department of Environmental Protection. (2006). *Developing Performance Based Measures Grant – Final Report.* Paper prepared for the U.S. Environmental Protection Agency.

Davanzo, T. [U.S. Environmental Protection Agency], personal communication, October 10, 2007.

Delmas, M.A. (2007). *Voluntary Agreements to Improve Environmental Quality: Are late joiners the free riders?* University of California, Santa Barbara: Institute for Social, Behavioral and Economic Research (ISBER Publications, Paper 07).

Dethman, L. (2004). *Final Report: 2003 Neighborhood Workshops Evaluation.* Seattle, WA: Dethman & Tangora LLC. Paper prepared for Seattle Public Utilities.

Environment Agency. *Greener Business is Good Business – Spotlight on business: Environmental Performance in 2003.* Bristol, UK: Author.

Evans, J.B.T. *Bias in Human Reasoning: Causes and Consequences* (1998), Hove, UK: Lawrence Erdbaum Associates.

Florida Department of Environmental Protection. *Planning, Integrating, and Supporting Priorities: Joint Planning to Enhance Environmental Results, Final Grant Report.* Paper prepared for U.S. Environmental Protection Agency.

Fornell, C. (2006). ASCI Commentary: Federal Government Scores. Retrieved November 29, 2007, from http://www.theacsi.org/index.php?option=com_content&task=view&id=162&Itemid=62.

Freedman, J.L. and Fraser, S.C (1996). "Compliance with Pressure: The Foot-in-the Door Technique," *Journal of Personality and Social Psychology* 4, no. 1 (1996): 195-202,

Gawande, A. (2007). *Better: A Surgeon's Notes on Performance.* New York: Metropolitan.

Gilmore Research and Social Marketing Services, Inc. (2006). *King County Environmental Behavior Index, Second Measurement: April/May 2006, Summary Report.* Paper prepared for King County Department of Natural Resources and Parks.

Gorman, J. (2006, May) *Region 2's Healthcare Compliance Initiative: Environmental Results From An Integrated Strategy* [PowerPoint Presentation]. Presented at the National Environmental Partnership Summit 2006, Atlanta, GA.

Government Accounting Standards Board. (2003). *Reporting Performance Information: Suggested Criteria for Effective Communication (Special Report).* Norwalk, CT: GASB.

Government Performance and Results Act of 1993 (GPRA).

Gray, W.B., & Shadbegian, R.J. (2005). Determinants of Environmental Compliance: Plant, Firm, and Enforcement Factors. Paper produced for the U.S. Environmental Protection Agency.

Gray, W.B., & Shadbegian, R.J. (2005). When and Why do Plants Comply? Paper Mills in the 1980s. *Law & Policy, 27*(2), 238-261.

Gray, W.B., & Shadbegian, R.J. *The Environmental Performance of Polluting Plants: A Spatial Analysis.* U.S. Environmental Protection Agency Grant #R-828824-01-0.

Gunningham, N., Kagan, R.A., & Thornton, D. (2004). Social License and Environmental Protection: Why Businesses Go Beyond Compliance. *Law & Social Inquiry,* 307-341.

Hatry, H.P. (1999). *Performance Measurement: Getting Results*. Washington, DC: Urban Institute.

Howard-Grenville, J., Nash, J., & Coglianese, C. (2007). *Constructing the License to Operate: Internal Factors and their Influence on Corporate Environmental Decisions*. Corporate Social Responsibility Initiative, Working Paper No. 27. Cambridge, MA: John F. Kennedy School of Government, Harvard University.

Industrial Ecology Program, School of the Environment, University of South Carolina. (2004). *USEPA/OECA/OC State and Tribal Assistance Grants (2000-2001), South Carolina: Measuring the Effectiveness of On-Site Inspection with Non-Regulatory Follow-Up by Industry Sectors, Final Report.* Retrieved September 13, 2007, from http://www.epa.gov/compliance/resources/reports/stag/fy2001/stag01-sc-onsite-final.pdf. Paper prepared for the U.S. Environmental Protection Agency.

Industrial Economics, Inc. (2003). *Evaluation of EPA New England's Colleges and Universities Initiative.* Cambridge, MA: Author. Paper prepared for the U.S. Environmental Protection Agency.

Industrial Economics, Inc. (2003). *Evaluation of EPA New England's Department of Public Works Initiative*. Cambridge, MA: Author. Paper prepared for the U.S. Environmental Protection Agency.

Jordan, A., Wurzel, R.K.W., & Zito, A. R. (Eds.) (2003). *New Instruments of Environmental Governance? National Experiences and Prospects*. London: Frank Cass.

Kahneman, D., & Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist, 39*(4), 341-350.

Kaplan, R & Norton, D. (2001). *The Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment,* Harvard Business School Press.

Kelman, S (2006). "Improving Service Delivery Performance in the United Kingdom: Organization Theory Perspectives on Central Intervention Strategies," *Joirnal of Comparative Policy Analysis,* 8:4, 393-419.

Kelman, S. (2005). *Unleashing Change: A Study of Organizational Renewal for Government*. Washington, DC: Brookings.

Kennedy School of Government (2001). *Get Results through Performance Management: An Open Memorando to Government Executives,* Harvard University.

Kettl, Donald F (1997). "The Global Revolution in Public Management: Driving Themes, Missing Links," *Journal of Public Policy and Management,* 16:3, 447-462.

Kotler, P., Roberto, N., & Lee, N. (2002). Social *Marketing: Improving the Quality of Life.* Thousand Oaks, CA: Sage.

Kowalewski, Rick (1996). "Using Outcome Information to Redirect Programs," U.S. Coast Guard Office of Marine Safety, Security, and Environmental Protection.

The Lewin Group. (2007). *Evaluaton of the District of Columbia's Child Support Co-Location Demonstration.* Paper prepared for the DC Child Support Service Division.

Locke, E. and Latham, G (2002). "Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey," *American Psychologist* 57-9, 705-717.

Lotspeich, K, & Fix, M. (2003). *E-Government and Regulation: The Department of Labor's Web-Based Compliance Assistance Resources.* Washington, DC: The Urban Institute. Paper prepared for the U.S. Department of Labor, Employment and Training Administration.

March, J.G. (1999). *The Pursuit of Organizational Intelligence.* Malden, MA: Blackwell.

Massachusetts Department of Agricultural Resources, Pesticide Bureau. *Development and Evaluation of an Integrated Pest Management Compliance Assistance Program to Protect the Health of Children in Schools, Daycare Center and Child Care Programs.* Paper prepared for the U.S. Environmental Protection Agency.

Massachusetts Department of Environmental Protection. *Final Report on Cooperative Agreement: CX 82949001.* Paper prepared for the U.S. Environmental Protection Agency.

Metzenbaum, S.H. (2005). Strategies for Using State Information: Measuring and Improving Program Performance. In J.M. Kamensky & A. Morales (Eds.), *Managing for Results 2005* (pp. 277-347). Washington, DC: IBM Center for the Business of Government.

Metzenbaum, S.H. (2006). *Performance Accountability: The Five Building Blocks and Six Essential Practices.* Washington, DC: IBM Center for the Business of Government.

Metzenbaum, S.H. (Spring 2007). "Can Environmental Agencies Make Better Use of Accident, Incident, and Inspection Information?", *ECOStates*, Washington, DC: ECOS.

Metzenbaum, S.H. et al (2007). *Memo on Measurement for Environmental Managers: Recommendations and Reference Manual.* College Park, MD: Environmental Compliance Consortium.

Michigan Combined Sewer Overflow and Sanitary Sewer Overflow 2005 Annual Report, http://www.deq.state.mi.us/documents/deq-wb-csossoreport05.pdf

Missouri Department of Natural Resources. (2002). *USEPA/OECA/OC State and Tribal Assistance Grants (1999-2000), Missouri: Measuring Performance from Enforcement and Compliance Assurance in Missouri Final Report.* Paper prepared for the U.S. Environmental Protection Agency.

Morris, L.A., Mazis, M.B., & Barofsky, I. (1980). *Product Labeling and Health Risks*, Banbury Report No. 6. Cold Spring Harbor Laboratory.

National Highway Traffic Safety Administration. *Keeping it Clicking in California*: *A Case Study of Sustained Success with a High Seat Belt Use Rate*. Retrieved September 30, 2007, from http://www.nhtsa.dot.gov/people/injury/airbags/buckleplan/bua_website/Archive-04/Cases/California.html.

New Mexico Environment Department. *Final Report, EPA Enforcement and Compliance Assurance Grant, Class V UIC Compliance Assistance*. Paper prepared for the U.S. Environmental Protection Agency.

Ohio EPA, Division of Air Pollution Control. (2004). *Quarterly Reporting Form for EPA's 2000 Enforcement and Compliance Assurance Grant, RMP/TRI Regulatory Awareness - Final Report*. Columbus, Ohio: Author. Paper prepared for the U.S. Environmental Protection Agency.

Office of Management and Budget (2006). Guide to the Program Assessment Rating Tool. Washington, DC: Author.

Penner, B.A. (2006). *Park Heights Project: Final Report*. Paper prepared for the Maryland Department of the Environment.

Quinn Patton, M. (1997). *Utilization-Focused Evaluation: The New Century Text.* Thousand Oaks, CA: Sage.

Rogers, E.M. (1995). *Diffusion of Innovations (Fourth Edition).* New York: The Free Press.

Runyon, C.W. (2003). "Introduction: Back to the Future – Revisiting Haddon's Conceptualization of Injury Epidemiology and Prevention," *Epidemiologic Reviews, 25,* 60-64.

Scholz, J.T. (1984). Cooperation, Deterrence, and the Ecology of Regulatory Enforcement. *Law and Society Review, 18*(2), 179-224.

Scott, Graham et al (1997). "New Zealand's Public Sector Management Reform: Implicatons for the United States," *Journal of Public Policy and Management* 16:3, 356-381.

Shapiro, A., & Dain, A. *Inside Perspectives: An Analysis of EPA Region 1's College and DPW Self-Audit Initiatives*. Cambridge, MA: John F. Kennedy School of Government, Harvard University. Paper prepared for U.S. Environmental Protection Agency, Region 1.

Shewmake, T. (2003). Calculating and Communicating Environmental Compliance Rates. *The Journal of the Environmental Council of the States,* 23-26.

Shewmake, T. (2004). Using Compliance Rates to Manage. *The Journal of the Environmental Council of the States*, 17-22.

Shimshack, J.P., & Ward, M.B. (2006). Enforcement and Over-Compliance. *Journal of Environmental Economics and Management* (forthcoming).

Shimshack, J.P., & Ward, M.B. (2005). Regulator Reputation, Enforcement, and Environmental Compliance. *Journal of Environmental Economics and Management, 50*, 519-540.

Short, J.L., & Toffel, M.W. (2005). Coerced Confessions: Self-Policing in the Shadow of the Regulator. *Journal of Law, Economics, and Organization*, pp. 1-27.

Stalfort, D.C. & McHenry, W.(1994). "Achieving Results: A Guide to Implementing the Marine Safety Business Plan, U.S. Coast Guard Marine Safety Branch Reserve Training Center.

Stowers, G.N.L. (2004). *Measuring the Performance of E-Government*. Washington, DC: IBM Center for the Business of Government.

Tetra Tech EM Inc. (2004). *COMET Project Task 1 Technical Memorandum*. Denver, CO: Author. Retrieved September 24, 2007, from http://www.epa.gov/compliance/resources/reports/stag/fy2001/stag01-co-comet-final.pdf. Paper prepared for the Colorado Department of Public Health and Environment.

Tetra Tech EM Inc. (2003). *The "COMPASS" Project: A Compliance Assistance Project*. Denver, CO: Author. Paper prepared for the Colorado Department of Public Health and Environment.

Texas Commission on Environmental Quality. *EPA Cooperative Agreement Grant for Development of Enhanced Performance Measures for Enforcement and Compliance Assurance Programs Final Report*. Paper prepared for the U.S. Environmental Protection Agency.

Thornton, D., Gunningham, N., Kagan, R.A. (2005). General Deterrence and Corporate Environmental Behavior. *Law & Policy, 27*(2)*,* 262-288.

Toffel, M.W., & Short, J.L. (2007). *Coming Clean…and Cleaning Up? Examining the Effects of Self-Policing.* In preparation.

Tyler, T.R. (1991). Procedural Fairness and Compliance with the Law. *Swiss Journal of Economics and Statistics, 133* (2/2), 219-240.

U.S. Environmental Protection Agency. (2002). *Washington, Final Project Report: Analysis of Change in Generator Compliance Using Regulatory Compliance Indicators.* Washington, DC: Author.

U.S. Environmental Protection Agency. *U.S. Environmental Protection Agency Statistically Valid Non-Compliance Rates Draft Report*. Washington, DC: Author.

Urban Institute (2003), *Key Steps in Outcome Management.* Washington, D.C.

Varshavsky, T. (2003). *Media Drug Prevention and Public Service Advertising: Evaluating the National Youth Anti-Drug Media Campaign*. Unpublished senior thesis, Tufts University.

Washington Department of Ecology, Hazardous Waste and Toxics Reduction Program. (2002). *USEPA/OECA/OC State and Tribal Assistance Grants (1999-2000), Washington: Analysis of Change in Generator Compliance Using Regulatory Compliance Indicators Final Report*. Paper prepared for the U.S. Environmental Protection Agency.

Wogalter, M.S. (2006). *The Handbook of Warnings.* Mahwah, NJ: Lawrence Erlbaum.

Wogalter, M.S., Dejoy, D.M., & Laughery K.R. (Eds.) (1999).*Warnings and Risk Communication*. London: Taylor & Francis.

APPENDIX A: FEDERAL BENCHMARK EXAMPLES

The following table uses agencies' and programs' Program Assessment Rating Tool (PART) reviews to assess their measurement strategies in order to gather information that may assist EPA in developing its compliance assistance measures. Agencies and programs were selected for the review based on their use of intermediate outcome indicators, such as improved understanding or adoption of behavioral practices, as well as the specificity of their measures.

The rows of each table, along with an explanation of the data they contain, are as follows:
- Agency/Program and Relevance to EPA Compliance Assistance
    - This row provides the title of the agency or program being reviewed and the relevance of their activities to EPA's compliance assistance objectives.
- Relevant Performance Goal/Objective from Strategic Plans:
    - This row includes all goals from agency strategic plans that encompass compliance assistance, whether focused solely on compliance assistance or including compliance assistance with other tools.
- PART Targets and Measures Related to Compliance Assistance
    - This row lists all measures included in the PART review that relate to the agency's compliance assistance objectives, whether directly or indirectly. For example, measures related to the number or percentage of unwanted events were included, as a decrease in unwanted events could be related to successful compliance assistance projects.
- Methodology Agency Uses to Assess Metrics
    - This row includes the methodology used to collect and analyze data presented in the PART review. Methodology information was gathered from PART reviews, Strategic Plans, and Annual Reports.
- Authors' Assessment of Pros/Cons of Metrics and Measurement Approach
    - In this section, the authors offer their evaluation of the metrics and methodology used to gather information on the agency or program. Authors evaluated the feasibility, coherence, motivational value, and replication value of the measures, as well as whether the information given allows for benefits seen in the measures to be attributed to the agency's actions or federal investment.
- Implications/ Ideas for EPA
    - In this section, the authors highlight what EPA can learn from the actions taken by the agency or program being reviewed.

Agencies and programs included are (in alphabetical order):

- Department of Health and Human Services – Food and Drug Administration
- Department of Homeland Security - Coast Guard – Marine Environmental Protection
- Department of Labor – Mine Safety and Health Administration

- Department of Labor – Occupational Health and Safety Administration
- Department of Transportation - National Highway Traffic Safety Administration - Operations and Research Program
- Internal Revenue Service -- Taxpayer Service Assessment
- Small Business Administration – Small Business Development Centers

Agency /Program and Relevance to EPA Compliance Assistance	**Department of Health and Human Services - Food and Drug Administration** FDA provides compliance assistance for companies whose products are regulated by the administration.
Relevant Performance Goal/Objective from Strategic Plans	The FDA Strategic Action Plan lists a few compliance assistance related objectives under Strategic Goal 4, "Improve the Quality and Safety of Manufactured Products and the Supply Chain." [64] Objective 4.1 relates to the creation and modernization of science-based standards and tools in order to promote compliance with FDA regulations. To this end, FDA conducts outreach and education activities for retail food stores, drug product manufacturers, and biological products manufacturers. Under Objective 4.2, FDA attempts to prevent unsafe food or drugs from coming into contact with consumers by improving its inspection and compliance protocols. [65]
PART Targets and Measures Related to Compliance Assistance	**PART Targets and Measures for Food and Drug Administration Assessment: [66]** *Outcome Measures* **Reduce medication errors in hospitals through increased adoption of bar code medication administration technology.** *Explanation:* This measure tracks the adoption rate of bar code medication administration technology in hospitals. Evidence shows that this technology reduces medication administration errors. **Increase by 10 percent the percentage of American consumers who correctly identify that saturated fat increases the risk of heart disease.** (Baseline FY 2005 developed. Target year for accomplishment, FY 2007.) Target: by 2007, 81% of all Americans will correctly identify that saturated fat increases the risk of heart disease. *Explanation:* This measure tracks the percentage of consumers who can correctly identify that saturated fat increases the risk of heart disease. **Improve by 10 percent the percentage of American consumers who correctly identify that omega-3 fat is a possible factor in reducing the risk of heart disease.** (Baseline FY 2005 developed. Target year for accomplishment FY 2007.) Target: by 2007, 36% of all American consumers will be able to identify that omega-3 fat is a possible factor in reducing the risk of heart disease. *Explanation:* This measure tracks the percentage of consumers who can correctly identify that omega-3 fat reduces the risk of heart disease. **Increase by 40 percent the percentage of American consumers who correctly identify that trans fat increases the risk of heart disease.** (Baseline FY 2005 developed. Target year for accomplishment, FY 2007) Target: by 2007, 45% of all American consumers will be able to identify that trans fat increases the risk of heart disease.

[64] FDA Strategic Action Plan, http://www.fda.gov/ope/stratplan07/stratplan07.htm#Obj4_2

[65] FDA Strategic Action Plan, http://www.fda.gov/ope/stratplan07/stratplan07.htm#Obj4_2

[66] PART Review for Food and Drug Administration, http://www.whitehouse.gov/omb/expectmore/detail/10001057.2003.html

	Explanation: This measure tracks the percentage of consumers who can correctly identify that trans fat increases the risk of heart disease. *Output Measures* **Inspect blood banks and biologics manufacturing establishments each year.** Target: by 2005, 50% of all blood banks and biologics manufacturing establishments will be inspected per year. **Inspect medical device manufacturing establishments each year.** Target: by 2005, 20% of medical device manufacturing establishments will be inspected per year.
Methodology Agency Uses to Assess Metrics	Data for the outcome measures was self-reported, and came from regulatory agencies participating in the Voluntary National Retail Food Regulatory Program Standards. FDA staff compiled the results.[67] The results for the output measures are a count of all inspections conducted by FDA at blood banks and medical device manufacturing establishments, respectively, each year. The blood bank data comes from CBER's Regulatory Management System and Field Data Systems.[68] The results are displayed as percentages. For blood banks and biologics manufacturing establishments, inspections include determining whether establishments recover tissue from donor properly, test and screen tissue donors properly, and store and transport tissues properly. The count of medical device manufacturing establishments includes inspections that were either completed by FDA or through state contracts or partnership agreements.[69] The inspections, according to FDA 2006 Foods Performance Goals, were limited to Class II and Class III domestic and foreign medical device manufacturers, because they pose the greatest risk to the public. [70]
Authors' Assessment of Pros/Cons of Metrics and Measurement Approach	**Feasibility of measuring accurately and credibly:** The output measures' results were calculated from a count of inspections conducted by FDA directly or through partnerships, so the data source is reliable. The outcome measures were calculated based on self-reported data, which introduces the possibility of error. **Motivational value:** The metrics are motivational, in that inspections are targeted toward the highest risk blood banks and medical device manufacturers. Staff time was used in the most effective way, and the knowledge that the facilities being inspected were the highest risk facilities highlights the importance of conducting thorough inspections. The measures successfully linked strategic goals (preventing harm, educating the public about health) to staff actions (inspections). **Replication value:** This approach can be easily replicated, as it only requires that an agency or program

---

[67] FY 2007 Budget, Foods Performance Goals,
http://www.fda.gov/oc/oms/ofm/budget/2007/HTML/1aFoodsPerfGoalApp.htm
[68] FY Budget 2007, Biologics,
http://www.fda.gov/oc/oms/ofm/budget/2007/HTML/1cBiologicsPerfGoalApp.htm
[69] FY Budget 2007, Medical Devices,
http://www.fda.gov/oc/oms/ofm/budget/2007/HTML/1eMedicalDevicesPerfGoalApp.htm
[70] FY Budget 2007, Medical Devices,
http://www.fda.gov/oc/oms/ofm/budget/2007/HTML/1eMedicalDevicesPerfGoalApp.htm

	keep track of the number of inspections of each type of facility it has completed.
	Coherence: The measures used reflect the goals outlined in FDA's Strategic Plan. As such they are clear, coherent ways to evaluate FDA's progress toward realizing its objectives. **Diagnostic value:** Inspections allow FDA staff to determine which regulations facilities' are out of compliance with, which allows them to pinpoint problems and address them. The outcome measures related to American consumers' knowledge about heart disease offers less diagnostic information. There are numerous reasons why American consumers may or may not know the link between saturated fat, trans fat, or omega-3 fats and heart disease. **Attribution:** The information gathered does not lend itself to determining whether benefits are attributed to FDA actions or to federal investment.
Implications/ Ideas for EPA	FDA targets its inspection efforts toward the highest risk facilities, as MSHA is trying to do, which could be an effective strategy for EPA to employ in its compliance assistance programs. Inviting the most noncompliant facilities to participate in compliance assistance activities, and providing incentives, may result in greater environmental protection and regulatory compliance.

Agency /Program and Relevance to EPA Compliance Assistance	**Department of Homeland Security - Coast Guard – Marine Environmental Protection (MEP) Program** The MEP Program prevents oil and hazardous materials from entering navigable waters. If the oil or hazardous materials do enter the water, the MEP Program seeks to remove them. Coast Guard activities prevent harm to the environment through such actions as regulatory and policy development, boardings and inspections of vessels and facilities to ensure compliance, education, navigational positioning, and communications. Core activities for pollution prevention include (among others) partnerships with industry associations and education and outreach programs.[71]
Relevant Performance Goal/Objective from Strategic Plans	The Strategic Plan for the Department of Homeland Security (DHS) does not indicate the Department has goals or performance measures specifically related to compliance assistance. [72] However, the annual Performance and Accountability Report[73] does list a specific measure related to the Marine Environmental Protection Program, namely, the five-year average number of U.S. Coast Guard investigated oil spills greater than 100 gallons and chemical discharges into the navigable waters of the U.S. per 100 million short tons of chemical and oil products shipped in U.S. waters. According to the Performance and Accountability Report, "This measure evaluates how well the Coast Guard prevents discharges of chemicals or oil into U.S. navigable

---

[71] *U. S. Coast Guard Marine Safety and Environmental Protection Business Plan FY2001-2005*, United States Coast Guard Marine Safety And Environmental Protection Directorate Washington, D.C. August 2000, http://www.uscg.mil/hq/g-m/nmc/gendoc/fy2001pp.pdf

[72] Department of Homeland Security Strategic Plan — Securing Our Homeland, http://www.dhs.gov/xabout/strategicplan/,

[73] Department of Homeland Security Performance and Accountability Report FY 2006, http://www.dhs.gov/xlibrary/assets/cfo_par2006_fullreport.pdf

	waters by comparing the current period to those of previous periods. Information recorded in the Coast Guard's Marine Information for Safety and Law Enforcement database is generally complete when the database is accessed. Some incidents are never reported, however, and some information is delayed in reaching the Coast Guard. Performance data will be revised as U.S. Army Corps shipping volume data becomes available. Duplicate information may occasionally be entered or an incident inadvertently omitted or incorrectly coded. Formal verification procedures strive to rectify any errors, and program logic and comprehensive user guides have been developed to ensure that data is highly reliable. The revised performance data will be available at the end of FY07 and available in next year's Performance and Accountability Report."
PART Targets and Measures Related to Compliance Assistance	**PART MEASURES:** [74] *Outcome Measure:* **Five-year average number of chemical discharge incidents and oil spills greater than 100 gallons per 100 million tons shipped** Target: by 2006, achieve a five-year average of 19 chemical discharge incidents and oil spills greater than 100 gallons per 100 million tons shipped. The program has achieved declining number for this indicator for every year shown in the PART assessment, starting from a high of 47.9 in 1999 to a value of 16.3 in 2006. The program has outpaced its target for every year shown in the PART assessment. *Explanation:* This performance measure indicates the five-year average number of USCG investigated incidents involving the discharge of chemicals or oil (more than 100 gallons) into navigable waters of the U.S. per 100 million short tons of chemicals and oil products shipped in U.S. waters. Only discharge incidents from maritime sources into U.S. waters are counted. *Efficiency Measure:* **Ratio of prior period to current period: Five-year average number of Oil Spills (>100 gal)and Chemical discharges, per 100M short tons shipped / Ratio of the current period to the prior period 5-year average Operating Expense Authority for Marine Environmental Protection** Target: by 2008, achieve a ratio of 0.97 for the Five-year average number of Oil Spills (>100 gal) and Chemical discharges, per 100M short tons shipped / Ratio of the current period to the prior period 5-year average Operating Expense Authority for Marine Environmental Protection. The value for this indicator in 2006 was 1.03, compared to a target of 0.95. No historical data are shown. Note: at the time the program underwent a PART assessment in 2003, the program was reporting two performance measures: **amount of oil spilled per million gallons shipped**, and **marine debris per mile of shoreline surveyed**. For internal agency reporting, Coast Guard also tracked the total number of oil and chemical spills, while the Port State Control program reported the number of foreign-vessel pollution ticket cases. In its 2003 Performance Accountability Report, the agency stated that, "The volume of oil spilled per million gallons shipped, the number of debris items collected per mile of shoreline surveyed and the total number of pollution incidents are lagging indicators that may not reflect current USCG Marine Environmental Protection performance. The Coast Guard is conducting a mission analysis project that will study both leading and lagging indicators including upper and lower control limits that may lead to better performance measures." The agency ultimately changed these measures

---

[74] Coast Guard Marine Environmental Protection PART Review,
http://www.whitehouse.gov/omb/expectmore/detail/10001073.2003.html

	to the current measures shown above.
Methodology Agency Uses to Assess Metrics	The Coast Guard's outcome measure (a five-year average of current and four previous years' number of chemical spills, and oil spills greater than 100 gallons, discharged into U.S. navigable waters per 100 million short tons of chemicals and oil products shipped) evaluates how well the Coast Guard prevents discharges of chemicals or oil into U.S. navigable waters by comparing the current period to those of previous periods. A five-year average is used to dampen the impact of year-to-year variation and to ensure that trends are apparent. Only discharge incidents from maritime sources into U.S. waters are counted. Discharges onto land, into the air, or into enclosed spaces are excluded, as are discharges from non-maritime sources. Discharges from naval and other public vessels; fixed platforms and pipelines, and discharges from unspecified, unclassified, and unknown sources are also excluded. Data are collected from USCG Marine Information for Safety and Law Enforcement System. Note that Coast Guard keeps detailed records about the characteristics of pollution incidents. Although these indicators are not reflected in the national PART or GPRA measures, the agency does use these data for internal tracking and program management. For example, the Coast Guard has reports on oil spills dating back from 1973 to 2004, and is able to show trends in the total number and volume of oil spills by spill size and the number of oil spills over 1,000 gallons. Since passage of the Oil Pollution Act of 1990, the Coast Guard has tracked number and volume of spills by: specific waterbody, location, source (e.g., pipeline, tankbarge, etc.), type of oil, and Coast Guard District.[75]
Authors' Assessment of Pros/Cons of Metrics and Measurement Approach	**Feasibility of measuring accurately and credibly:** Coast Guard relies on a well-established data collection system (Marine Information for Safety and Law Enforcement System), which appears to record reports by Coast Guard staff (however, whether Coast Guard employees are the original source of the data has not been confirmed). Having a centralized data system and using similar methodologies and assumptions in recording data make it more accurate and credible. **Motivational value:** This measure is motivating, in the sense that it is very specific and makes clear what the Coast Guard's priorities are. Coast Guard has stated that it is important to allow field staff some operational flexibility in adjusting their strategies to meet the targets,[76] and this helps the measure be more effective in motivating staff (since employees not only have a goal, but some say over the strategies they can use to reach it). On the other hand, using a five-year average, while helping the measure's coherence, may make it somewhat less motivating (e.g., if there was a large spike in spills and discharges last year, it may be somewhat de-motivating to staff to know that whatever progress they make this year, the five-year average measure will still reflect the previous year's spike.) **Replication value:** The specificity of the Coast Guard's goal makes it replicable for other agencies dealing with chemical discharges and oil spills. However, other agencies may not have the data collection infrastructure in place to collect data of a similar quality and specificity. **Coherence:** The current Coast Guard outcome measure is very specific yet somewhat complex, and

[75] U.S. Coast Guard, Pollution Incidents In and Around U.S. Waters A Spill/Release Compendium: 1969 – 2004, http://www.uscg.mil/hq/gm/nmc/response/stats/aa.htm

[76] Kowalewski, Rick, Commander, US Coast Guard, Using Outcome Information to Redirect Programs: A Case Study of the Coast Guard's Pilot Project Under the Government Performance and Results Act, April 1996.

therefore may be slightly difficult for those outside the agency to understand. For example, those outside the agency may not know that 100 gallons is generally a considered a very small oil spill, and so by only looking at oil spills of greater than 100 gallons, the Coast Guard is essentially eliminating only the smallest spills from its measurement. Spills of 100 gallons or less constitute the vast majority of the number of oil spills, but spills up to 1,000 gallons have historically been responsible for a relatively small proportion of the volume of oil spilled (at least until 1996, when Coast Guard succeeded in dramatically the volume of oil released from spills over 100,000 gallons).[77] By counting the number of discharge incidents and spills, but not the volume, in its current PART measure, Coast Guard does not fully show its success in reducing the number of oil spills. Fortunately, the agency does track this information internally, even if it is not included in its PART measure. Also note that the Coast Guard clearly made a choice after 2003 to from a volume-based metric to a count-based metric. Although the agency said that its intent was to use a leading, rather than a lagging, indicator, it is not clear to this author that they have succeeded in doing so. Further conversation with Coast Guard may shed light on this issue.

Diagnostic value:
While the Coast Guard's PART measure is of limited diagnostic value, the more detailed data that the agency tracks (e.g., number and volume of spills by specific waterbody, location, source, type of oil, and Coast Guard District) are of great value in diagnosing problems and pinpointing what strategies are working. For example, Coast Guard data show a dramatic drop in the volume of oil discharged from large spills in 1991, and another substantial drop in 1997. Correlating these changes with changes in Coast Guard strategies or other external events (such as passage of the Oil Pollution Act of 1990) could be very helpful in understanding what strategies are working.

Attribution:
The Coast Guard's PART measure itself does not lend itself to discerning what changes in performance may be attributable to the agency's own actions vs. other agencies' efforts or external events. The PART measure also does not explicitly differentiate outcomes associated with compliance assistance vs. enforcement or other tools the agency uses to reduce chemical discharges and oil spills. However the more detailed data the agency tracks should help Coast Guard with these attribution issues. For example, if the agency observes changes in performance that occur at the same time, in the same place, or affecting the same sources as specific agency initiatives, then a reasonable case could be made that those agency initiatives led to the observed outcomes.

Implications/ Ideas for EPA	A key implication from the Coast Guard's performance measures is that other agencies that do conduct compliance assistance as part of their missions do not necessarily measure these efforts distinctly, but instead may measure the overall impact of compliance assistance, enforcement, and other efforts on specific desired outcomes. Collecting more detailed data (e.g., tracking outcomes for specific target audiences, specific geographic regions, or specific types of pollution) could help EPA understand whether their efforts are having the desired effect. For example, with regard to EPA's measure "12 percent reduce, treat, or eliminate pollution;" the agency could keep a measure similar to this at the national level, but have supporting data that allows EPA to break this measure down into specific industries, types of pollution, and geographic location. Then, by tracking these specific measures over time, EPA could determine if a significant change in outcomes (e.g., a jump in the percentage of entities reducing, treating, or eliminating pollution in a specific sector and a specific geographic area) coincided with a particular compliance assistance effort. While not proving that the compliance assistance led to the change in observed outcome, this would be a much stronger argument if EPA had available the specific data by sector and geographic

---

[77] *Ibid.*

area.

The same approach to tracking disaggregated data could also help EPA diagnose which compliance assistance efforts are most effective in reaching the desired goal. For example, this type of data would lend itself well to tracking the results of controlled experiments (e.g., trying different compliance assistance approaches in different parts of the country, and then observing which regions had the greatest changes in outcomes.)

The Coast Guard example also shows the value of using standard, consistent performance measurement methods over time. For example, the fact that Coast Guard is able to track trends in some measures since 1973, and other measures since 1991, is very helpful in showing the agency's long-term effectiveness. Once EPA chooses a measure that it finds acceptable, the agency may be well-served by keeping that measure (even if it is not perfect), so that it can be used for long-term trend analysis. The Coast Guard experience also suggests that centralized data collection (ideally by agency staff) is useful in establishing accurate, credible data. To the extent possible, it is therefore useful to have agency staff verify changes in practices (rather than relying on self-reported data). On the other hand, for practical reasons such as resource constraints EPA may be forced to rely on self-reported data.

Agency /Program and Relevance to EPA Compliance Assistance	**Department of Labor - Mine Safety and Health Administration (MSHA)[78]** The Mine Safety and Health Administration provides compliance assistance for mine operators in accordance with the Federal Mine Safety and Health Act of 1977 and the Mine Improvement and New Emergency Response (MINER) Act of 2006. Within initiatives such as the Cooperative Accident Reduction Effort (CARE), the administration carries out enforcement and compliance assistance actions at high-risk mines. It also provides technical assistance and support to mine operators so that they can improve their training programs and effectively address health and safety problems. The administration supports states' efforts to provide compliance assistance by issuing grants to state education and training programs.
Relevant Performance Goal/Objective from Strategic Plans	The Department of Labor Strategic Plan includes compliance assistance performance goals. Under the 3rd Strategic Goal, "Safe & Secure Workplaces," DOL plans to improve health and safety in the workplace through compliance assistance and the enforcement of health and safety regulations (Performance Goal 3A).[79] One performance goal listed under "Safe & Secure Workplaces" pertains directly to mines; DOL seeks to reduce fatalities, injuries, and illnesses suffered by mine employees at the workplace (Performance Goal 3B).[80] This performance goal does not explicitly mention compliance assistance, but compliance assistance plays an indirect role in achieving this goal. Under this goal, DOL conducts outreach through its own programs such as the Preventive Roof/Rib Outreach Program, Winter Alert, and Mine Emergency Preparedness. It also funds state programs that offer safety training to mines.[81]

---

[78] Mine Safety and Health Administration Fact Sheet, http://www.msha.gov/MSHAINFO/MSHA_Informational_Flyer.pdf
[79] DOL Strategic Plan for Fiscal Years 2006-2011, http://www.dol.gov/_sec/stratplan/strat_plan_2006-2011.pdf.
[80] DOL Strategic Plan for Fiscal Years 2006-2011, http://www.dol.gov/_sec/stratplan/strat_plan_2006-2011.pdf.
[81] DOL Strategic Plan for Fiscal Years 2006-2011, http://www.dol.gov/_sec/stratplan/strat_plan_2006-2011.pdf.

PART Targets and Measures Related to Compliance Assistance	PART Measures for Mine Safety and Health Administration Assessment[82]
	Outcome Measures
	Fatal injury incidence rate.
	Baseline is FY 2003 rate of .0219 fatalities per 200,000 hours worked by mine employees. (Long-term)
	Target*:* by 2007, there will be 0.0201 fatalities per 200,000 hours worked by mine employees.
	Explanation: This measure tracks MSHA's annual performance in reducing the fatal injury incidence rate (per 200,000 hours worked by mine employees) by a total of 15% from the baseline to FY 2008.
	All-injury incidence rate.
	Baseline is FY 2000 rate of 5.07 all-injuries per 200,000 hours worked by mine employees. (Long-term)
	Target*:* by 2007, there will be 2.82 injuries per 200,000 hours worked by mine employees.
	Explanation: This measure tracks MSHA's annual performance in reducing the all-injury incidence rate (per 200,000 hours worked by mine employees) by a total of 50% from the baseline to FY 2008.
	Output Measures
	Reduce respirable coal dust samples exceeding applicable standards by 5% for designated occupations.
	Target*:* by 2007, 9.00% of all respirable coal dust samples will exceed the applicable standards.
	Explanation: This measure tracks MSHA's annual performance in reducing the percentage of respirable coal dust samples exceeding the applicable standards. Samples taken by mine inspectors.
	Silica dust samples exceeding applicable standards by 5% for designated high risk occupations.
	Baseline = 9% of samples out of compliance in FY 2002. Future goals are being revised based on exceeding FY 2003 target.
	Target*:* by 2007, 17.90% of all silica dust samples will exceed applicable standards by 5% for designated high risk occupations.
	Explanation: This measure tracks MSHA's annual performance in reducing the percentage of silica dust samples exceeding the applicable standards. Samples taken by mine inspectors.
	Noise exposures above the citation level by 5% for designated high risk occupations.
	Baseline = 9.3% of samples out of compliance for FY 2000 - FY 2001.
	Target*:* 23.00% of all samples will be above the citation level by 5% for designated high risk occupations.
	Explanation: This measure tracks MSHA's annual performance in reducing the percentage of noise exposures in all mines. Samples taken by mine inspectors.
	Reduce noise exposures above the citation level in coal mines by 10%.
	Target: by 2007, 4.80% of noise exposures will be above the citation level in coal mines by 10%.
Methodology Agency Uses to Assess Metrics	The outcome measures (fatal injury incidence rate and all-injury incidence rate) relate to MSHA's goal of reducing work-related fatalities and injuries. Mine operators report all work-related injuries, accidents, and illnesses to the MSHA within ten days; all

[82] PART Review for Mine Safety and Health Administration,
http://www.whitehouse.gov/omb/expectmore/detail/10001101.2003.html

	fatalities are reported immediately. MSHA analyzes data using its management information system.[83] To calculate the fatality and all-injury rates, MSHA divides the number of fatalities and injuries times 200,000 hours worked by the actual hours employees work at mines. Mine operators provide MSHA with data on the number of hours employees work at mines. To ensure data quality, MSHA conducts data audits and employs built-in data checks during data entry. For the output measures, mine inspectors collected samples of coal dust, silica dust, and noise exposures. Under DOL's 2006 Strategic Plan, MSHA sampling protocols were changed to enable inspectors to target the highest-risk mines and occupations. The administration focused resources on occupations with greater exposure to noise and silica (i.e., where the percentage of silica dust and noise samples that are less than 50% of the permissible exposure limits had decreased).[84] At the time of the PART review, however, MSHA was not able to fully target its inspections toward hazardous mines because of Mine Act requirements.[85] MSHA inspectors, or safety and health compliance specialists, follow established procedures when collecting noise, silica dust, and coal dust samples.[86] Once the samples are collected, they are weighed by devices which automatically enter the results into the database. MSHA and the National Bureau of Standard created quality control program that periodically reviews data produced by the devices in order to ensure its consistency and accuracy.[87]
Authors' Assessment of Pros/Cons of Metrics and Measurement Approach	**Feasibility of measuring accurately:** Since mine operators regularly send fatality, injury, and illness data to MSHA, measurements are very accurate for outcome measures. The quality control program and frequent checks of the data further increase the accuracy of MSHA data. Enabling mine operators to report unwanted incidents directly to the administration provides MSHA with the most current data possible. Focusing inspection on the most hazardous workplaces is a successful strategy for decreasing the number of fatalities, injuries, and illnesses that employees suffer; however, sampling mines based on this criterion introduces bias. A random sampling of mines would produce a more accurate result. Even so, given MSHA's strong interest in minimizing fatalities, injuries, and illnesses and its limited resources, conducting inspections in this manner produces the most useful data for its purposes. **Motivational value:** The measurement approach is motivating because it provides MSHA safety and health specialists and mine operators with the most current data possible. Additionally, the approach clearly reflects MSHA's priorities with respect to miners' health and safety. By choosing to focus on the most hazardous mines rather than taking a random sampling of all mines, MSHA indicates that its first priority is decreases in the fatality and injury rates even though the administration was restricted by the Mine Act. Data analysis and sharing also make this measurement approach effective and motivational. When mine operators report incidents to MSHA, the new data is incorporated into MSHA databases. Dust samples taken by inspectors are instantly entered into the databases. MSHA shares certain data on fatalities, accidents, and

[83] PART Review for Mine Safety and Health Administration,
http://www.whitehouse.gov/omb/expectmore/detail/10001101.2003.html
[84] DOL Strategic Plan for Fiscal Years 2006-2011, http://www.dol.gov/_sec/stratplan/strat_plan_2006-2011.pdf.
[85] PART Review for Mine Safety and Health Administration,
http://www.whitehouse.gov/omb/expectmore/detail/10001101.2003.html
[86] DOL Annual Report, Fiscal Year 2006, Performance and Accountability Report,
http://www.dol.gov/_sec/media/reports/annual2006/PDF/2006annualreport.pdf
[87] MSHA Strategic Plan FY03-08, http://www.msha.gov/MSHAINFO/STRAPLAN/STRAPLAN.pdf

injuries with the public, and maintains an Intranet website that shares information with other health and safety specialists. In addition, MSHA provides mine operators with a summary of pertinent data so that they may compare their mines with similar mines in their region or nationwide.[88] The widespread dissemination and use of the data can be motivational for mine operators and inspectors.

MSHA does not use performance measurement data to inform its decisions on budget requests. However, the MSHA does consider performance measurement data when rating the performance of managers and program partners, which serves as motivation to reach program goals and targets. [89]

Replication value:
This methods utilized in this review are transferable to other federal agencies that evaluate numerous facilities. Involving facility operators and managers in the evaluations process, while also utilizing inspectors for certain metrics, may be a successful strategy for other federal agencies. Certain aspects of MSHA's data collection methods (the quality control program and the device that analyzes samples and automatically enters results into the database) would require more established data collection programs than some agencies currently possess, such as the quality control program and the devices that analyze sample The quality control methods used are particularly replicable, as are the methods used for transferring sampling results to the database.

Coherence:
MSHA's outcome measures are clear and understandable.

Diagnostic value:
Some measures provide diagnostic value. Metrics with respect to air quality and noise exposure, for example, are helpful in determining the success or failure of MSHA's compliance and enforcement activities. The air quality and noise exposure measures directly relate to three health conditions (Black Lung Disease for coal dust, Silicosis for silica dust, and hearing loss for noise exposure), so a decrease in the amount of coal dust, for example, will most likely lead to a decrease in the Black Lung Disease incidence in the future and vice versa. However, metrics such as the fatal injury and all injury incidence rates may have multiple possible causes. For example, in 2006, 19 miners died in serious accidents at three mines: the Sago and Alma mines in West Virginia and the Darby mine in Kentucky. In the 3 years preceding those events, the fatality and injury rates had declined significantly, but the deaths at those three mines led to an increase in the fatality and injury rates in 2006.[90] In that case, the reversal of the fatal injury incidence rate trend correctly highlights the need for safety improvements in mines, but it does not show whether the increase in fatalities are spread out over the industry as a whole or represent events like those in the Sago, Alma, and Darby mines.

Attribution:
This data does not distinguish benefits caused by MSHA actions or by MSHA investment. The PART review does not include any measures that specifically

---

[88] MSHA Strategic Plan FY03-08, http://www.msha.gov/MSHAINFO/STRAPLAN/STRAPLAN.pdf, p. 17.
[89] PART Review for Mine Safety and Health Administration, http://www.whitehouse.gov/omb/expectmore/detail/10001101.2003.html
[90] DOL Annual Report, Fiscal Year 2006, Performance and Accountability Report, http://www.dol.gov/_sec/media/reports/annual2006/PDF/2006annualreport.pdf, p. 98.
[91] PART Review for Small Business Development Centers – Small Business Administration, http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html

	reference compliance assistance or enforcement actions; the measures could be the result of either of those two types of actions. However, MSHA awards grants to organizations that provide training to mine operators. MSHA requires that grantees meet specific goals and monitors grantees to ensure that progress is being made.[91] Data from the grant program could help in attributing success or failure to MSHA or its partners.
Implications/ Ideas for EPA	MSHA makes excellent use of information technology in order to ensure rapid data transmission and entry. The device that enables inspectors to automatically evaluate samples and enter data minimizes user error in data entry. The quality control program and data checks also serve to minimize user error in the entry of fatality and injury data. EPA could use similar techniques to ensure that its databases are up to date and accurate. Automatic entry of inspection information during compliance assistance projects would be beneficial in determining whether the compliance assistance project is positively impacting the compliance assistance recipients. Though it is currently hindered in part by the 1977 Mine Act, MSHA is attempting to target its sampling efforts toward the most hazardous worksites. This initiative is important in enabling MSHA to make more progress toward decreasing fatality and injury rates with its limited resources. EPA, similarly, may be able to target more of its compliance assistance toward the facilities that are most in need of compliance assistance. Actively seeking out the most noncompliant facilities, instead of relying on facilities to volunteer for compliance assistance projects, may produce higher compliance rates and greater environmental benefits.

Agency /Program and Relevance to EPA Compliance Assistance	**Department of Labor (DOL) – Occupational Health and Safety Administration (OSHA)** Similar to EPA, OSHA seeks to reach its goal to protect workers' safety and health through a combination of enforcement program, outreach, education, compliance assistance and voluntary cooperative programs. According to DOL's strategic plan, [92] the agency plans to, "continue to direct inspections and outreach at establishments and industries with the highest injury, illness, and fatality rates and will respond to complaints of serious workplace hazards. As part of the Department's outreach effort, selected sites with high injury and illness rates will be notified in writing of available services for addressing workplace hazards. Small business employers who receive notification will be provided an opportunity to seek assistance through the free, DOL-funded State Consultation Program. These efforts will be supplemented by National and Local Emphasis Programs designed to target unsafe conditions or high hazard industries. To complement its enforcement and standard-setting activities, the Department will provide compliance assistance, outreach, and training for employers and employees. DOL also offers a variety of cooperative programs including the Voluntary Protection Programs, the Alliance Program, the Strategic Partnership Program, the Consultation Program and its Safety and Health Achievement Recognition Program, under which employers, employees, and other stakeholders work with the Department to improve workplace safety and health."
Relevant Performance Goal/Objective from Strategic Plans	Strategic Plan Performance Goal 3A: Improve workplace safety and health through compliance assistance and enforcement of occupational safety and health regulations and standards Strategic Plan Performance Measures Related to CA Include: • Rate of workplace injuries and illnesses, measured by Calendar Year Bureau

---

[92] Department of Labor Strategic Plan for Fiscal Years 2006 – 2011,
http://www.dol.gov/_sec/stratplan/strat_plan_2006-2011.htm#ca

	of Labor Statistics Days Away Restricted and Transferred (DART) Rate for Private Sector • Rate of workplace fatalities
PART Targets and Measures Related to Compliance Assistance	**PART MEASURES:**[93] *Outcome Measures:* **Days away from work, restricted and transferred (DART) per 100 workers** Target: by 2011, achieve 2.0 DART per 100 workers (note, strategic plan indicates that this measure is tracked per calendar year by the Bureau of Labor Statistics, and that it only includes private sector employees) **Workplace fatalities per 100,000 workers for sectors covered by the OSH Act** (October-September data, as reported on January 1 of following year.) Target: by 2011, achieve a workplace fatality rate of 1.66 per 100,000 workers.
Methodology Agency Uses to Assess Metrics	OSHA has created two new long-term performance measures (1 and 2 under the performance measures section) in conjunction with the 2007 PART evaluation, which track slightly different illness, injury, and fatality data than the previous measures. The current measures are outcome-oriented and reflect key components of OSHA's mission. Beginning in 2007 the workplace injury and illness rate will be measured using the DART (days away from work, restricted work activity, and/or job transfer) rate rather than the previous DAFW (Days Away From Work) rate. The DART rate, which includes restricted work activity or job transfer, is a more comprehensive measure of injuries and illnesses which affect employees' work than the days away from work (DAFW) rate used in the FY 2003 – 2008 strategic plan. The rate of workplace fatalities will still be measured by the rate of OSHA-investigated deaths among workers covered by the OSH Act, but beginning in 2007 will be calculated several months after the conclusion of the performance period, allowing for a more complete count of fatalities. These performance measures will be analyzed both on a long-term basis (over the five years of the Plan) and on an annual basis (as measured through the Agency's annual operating plan and through the Department's Annual Performance and Accountability Report). [94] Bureau of Labor Statistics (BLS) occupational injury and illness numbers come from the BLS annual Survey of Occupational Injuries and Illnesses. The survey captures data from OSHA logs of workplace injuries and illnesses maintained by employers. The BLS Survey measures nonfatal injuries and illnesses only and excludes the self-employed; farms with fewer than 11 employees; private households; Federal government agencies; and, for national estimates, employees in State and local government agencies. Data for railroads and certain mining industries are not from the BLS survey, but are supplied to BLS from the Federal Railroad Administration and the Mine Safety and Health Administration.[95] The BLS Census of Fatal Occupational Injuries (CFOI is a Federal-State cooperative program that has been implemented in all 50 States and the District of Columbia since 1992. To compile counts of fatalities that are as complete as possible, the census uses multiple sources to identify, verify, and profile fatal worker injuries. Information about each workplace fatality--occupation and other worker characteristics, equipment involved, and circumstances of the event--is obtained by cross referencing the source

---

[93] PART Review for Occupational Safety and Health Administration,
http://www.whitehouse.gov/omb/expectmore/detail/10000336.2007.html
[94] PART Review for Occupational Safety and Health Administration,
http://www.whitehouse.gov/omb/expectmore/detail/10000336.2007.html
[95] U.S. Department of Labor, Bureau of Labor Statistics, Injuries, Illnesses, and Fatalities program, "People are asking...," (Frequently Asked Questions fact sheet), http://www.bls.gov/iif/peoplebox.htm#faqaa

	records, such as death certificates, workers' compensation reports, and Federal and State agency administrative reports. To ensure that fatalities are work-related, cases are substantiated with two or more independent source documents, or a source document and a follow-up questionnaire.[96] Note that BLS maintains detailed statistics about characteristics of occupational injuries, illnesses, and fatalities, even those these disaggregated data are not reported as part of the GPRA or PART measures. For example, BLS disaggregated its illness and injury data by type of industry, occupation, nature of injury/illness, number of hours worked before the event occurred, time/day of week of event, and worker demographics.[97] BLS maintains similar data for fatalities.[98]
Authors' Assessment of Pros/Cons of Metrics and Measurement Approach	**Feasibility of measuring accurately and credibly:** OSHA relies on BLS data for its PART measures, which distinguishes OSHA from other agencies reviewed. Involving additional agencies could introduce the possibility of error. Additionally, data is largely self-reported, so bias is a possibility if there are no mechanisms in place that encourage accurate reporting. **Motivational value:** The OSHA measures, and particularly the disaggregated data that BLS tracks (but does not report for GPRA and PART) are very specific, and this should make them motivating not only to agency staff, but also to others interested in affecting outcomes. For example, since BLS publicly reports its disaggregated data, compliance assistance providers can easily see which industries are responsible for the greatest numbers of illnesses and injuries, and this could help motivate them to offer their services to these industries. **Coherence:** The OSHA measures are relatively intuitive, although the details about which data are excluded (e.g., data from the self-employed, small farms, etc.) are important to interpreting the measures (since presumably the excluded data may differ systematically from the data captured in the measure.) However, since these exclusions are clearly stated and reported with the data, the measures are coherent. **Diagnostic value:** The PART measures themselves do not offer very much insight into potential problems, but the more specific disaggregated data that BLS tracks could be used to diagnose and address issues. **Attribution:** Disaggregated data may be effective in attributing benefits to either agency actions or federal investment. Since the data for the PART measures are less specific, they are less helpful in determining attribution.
Implications/ Ideas for EPA	

Agency /Program and	**Department of Transportation - National Highway Traffic Safety Administration (NHTSA) Operations and Research Program**

[96] U.S. Department of Labor, Bureau of Labor Statistics, Injuries, Illnesses, and Fatalities program, "Census of Fatal Occupational Injuries (CFOI)," http://www.bls.gov/iif/oshfat1.htm
[97] Bureau of Labor Statistics, Case and Demographic Characteristics for Work-related Injuries and Illnesses Involving Days Away From Work, http://www.bls.gov/iif/oshcdnew.htm#06b
[98] Bureau of Labor Statistics, Census of Fatal Occupational Injuries (CFOI) - Current and Revised Data, http://www.bls.gov/iif/oshcfoi1.htm

Relevance to EPA Compliance Assistance	NHTSA's compliance assistance program tries to prevent unwanted events, specifically fatalities. NHTSA also tries to promote environmentally sound practices, but does not have direct enforcement authority in the states as EPA does. Instead, NHTSA uses a combination of positive (grants) and negative incentives (grant withholding), assistance, tool development, marketing, data analysis, analysis sharing, and coaching to motivate states to adopt and enforce laws and practices it has identified that are effective in reducing fatality rates. Unlike EPA, NHTSA's organizational structure and measures integrate use of assistance and other tools.
Relevant Performance Goal/Objective from Strategic Plan	Enhance public health and safety by working toward the elimination of transportation-related deaths and injuries. 2011 target is 1.0 highway fatalities per 100 million vehicle miles traveled (VMT.)
PART Targets and Measures Related to Compliance Assistance	*Note: The "most future" date for which a target was set is noted below, but NHTSA sets an annual target for each measure.* *Long-Term Outcome Measures* **Highway fatalities per 1000 million vehicle miles traveled.** Target*:* By 2009, 1.35 highway fatalities per 1000 million vehicle miles traveled (VMT), down from a 1.7 fatality rate in 1996. *Long-Term/Annual Outcome Measures* **Passenger vehicle occupant highway fatality rate.** Target: By 2013, .90 passenger vehicle occupant highway fatality rate per 100 million passenger vehicle VMT compared to 1.25 in 2001. (Includes passenger cars, light trucks, vans, and SUVs.) *Annual Outcome Measures* **Non-occupant highway fatality rate.** Target: By 2013, 18 non-occupant highway fatality rate per 100 million VMT compared to .21 in 2001. (pedestrians, cyclists and occupants of motor vehicles not in transport and of non-motor vehicle transport devices) **Motorcycle rider highway fatalities.** Reduce the expected rate of increase in motorcycle rider highway fatalities per 100 million motorcycle VMT. Target: NHTSA has set a target rate of 62.00 fatalities per 100 million motorcycle miles traveled for 2007. (This is an ambitious target considering agency projections show an increase to 63.00 in 2007.) **Vehicle occupants using safety belts.** Target: By 2011, 83 percent of vehicle occupants using safety belts up from 73 percent in 2001. *Explanation:* 2005 & 2006 targets depend on States enacting and enforcing primary safety belt use laws **Restraint use among 0 to 7 year olds.** Target: By 2011, 88% restraint use among 0 to 7 year olds. (The agency changed its prior goal of reducing the number of child occupant fatalities, 0-4 years old, because the goal of 465 was surpassed in 2002, three years prior to the 2005 goal. NHTSA chose a new goal of increasing restraint use among 0 through 7. The agency re-baselined its restraint use target for 2007 after data showed a significant decline from 88 percent in 2002 to 82 percent in 2004 (data in 2003 was not collected and data for 2005 is not yet available). The agency set its past targets based off of this one data point, but with a second year of data now available it was better able to forecast and project future restraint use in setting out-year targets. NHTSA has set a new target of 85 percent for 2007.)

	High blood alcohol content. Target*:* By 2011, .45 fatality rate in high blood alcohol content (0.08+) crashes per 100 million Vehicle Miles Traveled. (*2003-2005 targets included all alcohol-related fatalities. (2006 target was revised to reflect that these drivers (.08 and above BAC) make up 85 percent of the alcohol problem.) *Efficiency Measures (Annual)* **Complete significant rulemaking actions**. Target*:* By 2011, 12 months to complete significant rulemaking actions. (Measure is restricted to time within the agency and does not include OST and/or OMB review periods) from 18 months in 2003. **Completion time for a defect investigation**. Average completion time for a defect investigation. Target: NHTSA will maintain the average completion time for a defect investigation at 8 months. (The Defects Investigation Program collects information, analyzes, and conducts investigations of potential vehicle safety defects …)
Methodology Agency Uses to Assess Metrics	*Fatality rate.* To count fatalities, NHTSA partially funds a Governor's safety advisor in every state to gather data about <u>every</u> traffic fatality and note key characteristics of the fatality. Those characteristics fall into nine general categories specified by the Haddon 3 x 3 matrix, developed using principles of injury epidemiology. ( ) For every fatality, NHTSA notes human, equipment, and environmental conditions (including location) before, during, and after each fatality as well as the time and date of the fatality. Thus, for example, NHTSA notes the age and sobriety of drivers and it notes post-accident costs. To normalize the number of fatalities to take into account the increasing number of cars on the road, NHTSA uses VMT data collected by FHWA, which also relies on states for primary data collection and then uses a variety of methods to adjust the heavily used VMT data.[99]. **Safety Belt Use.** To count safety belt use, NHTSA has developed an observer-based counting protocol, where observers stand at a number of selected street corners in every state and count belt use. This counting method is now done annually in every state so that NHTSA can see belt use rate trends nationally and at the state level. Although belt use count could be biased by the selection of observer location if people in different neighborhoods have different belt use patters, that bias is likely to be consistent over time so NHTSA can feel confident about the reliability of the trend.
Authors' Assessment of Pros/Cons of Metrics and Measurement Approach	**Feasibility of measuring accurately and credibly:** NHTSA's fatality counting system was started at the inception of the agency. It is not cheap, but the agency and Congress have consistently continued to fund and support in its maintenance and improvement because NHTSA's priorities and its daily work are heavily informed by the measures. **Motivational value:** NHTSA uses specific, challenging targets, which have been shown to have a strong motivating effect. It not only reports outcome indicators nationally, but also at the sub-national level in a way that corresponds to the responsibilities of distinct performance units (e.g., regions, states, localities.) This supports benchmarking among similar parties to look for those with better performance, and puts pressure on low performers through local press and legislative attention. **Replication:**

---

[99] www.fhwa.dot.gov/ohim/vm194.pdf   This technical document dates back to 1994, but provides an explanation of the technical measurement methods FHWA uses to calculate VMT.

NHTSA's diagnostic analysis sometimes points it to potentially promising practices. For example, when California's fatality rates dropped suddenly, NHTSA followed up to understand why and learned the state had adopted a primary enforcement law, authorizing police to stop and check for belt use without needing another reason to stop a vehicle. NHTSA's measurement method allowed it and other states to detect the "positive deviance" in CA compared to trends elsewhere. NHTSA then started to track and compare fatality rate changes in states adopting similar laws to changes in other states. Once it saw similar fatality rate reductions in the other states, NHTSA knew it had discovered a promising practice worth promoting.

Once it finds a promising practice, such as belt use primary enforcement laws and universal helmet laws, NHTSA tracks state uptake of those practices. It uses the uptake information to confirm the effectiveness of the interventions and to promote them in places where they have not yet been adopted.

Coherence:
NHTSA uses two types of outcome measures, fatality rate and behavioral change, that are easily understood by the public, by NHTSA staff, and by the many others whose actions can influence the outcomes. The fatality rate measure aligns with and supports US DOT priority goals.

NTHSA has set one overarching outcome target, reducing highway fatalities, and supports that with targets for four specific populations – passengers, non-occupants, motorcycle riders, and heavy alcohol users. Setting these specific targets communicate to NHTSA staff and the public where the agency is placing its priorities. It also reveals, by omission, that NHTSA is not able to focus on every problem that needs attention. It has not, for example, adopted a GPRA target to reduce accident rates of teen male drivers, although they are clearly a high risk population.

NHTSA has also adopted two intermediate outcome behavioral targets for practices known to reduce fatality rates – safety belt use and child restraint use. The behavioral targets are described in language that is easily understood by everyone. Moreover, because of NHTSA and insurance industry message campaigns such as "Seat Belt Use Saves Lives," the target audience understands why it is supposed to change its behavior.

Diagnostic value: Noting key characteristics in a consistent manner for a consistent set of categories (laid out by the Haddon matrix) and looking at performance associated with each characteristics allows NHTSA to conduct robust diagnostic analysis to find causal factors: preventable ones such as alcohol use, non-preventable ones such as age that are correlated with changes in the fatality rate, and promotion-worthy ones such a helmet use. NHTSA analyzes the information it collects to look for:

- Pattern variations in different subsets (e.g. higher fatality rates of young men and older drivers) that reveal possible problems and promising practices,
- Emerging problems (e.g., increasing motorcycle fatality rates following the elimination of motorcycle helmet laws),
- Relative importance of problems (e.g., which subsets of people/equipment/conditions have the highest fatality rates),
- Positive deviance (e.g., which subsets have the lowest fatality rates or

---

[100] Metzenbaum. "Strategies for Using State Information," pp. 32-33.
[101] Mark G. Solomon et al, "Evaluation of Click It or Ticket Model Programs," Technical Report DOT HS 809 498 prepared for the National Highway Traffic and Safety Administration, September 2002. Available at: http://www.nhtsa.dot.gov/people/injury/airbags/clickit_ticket/clickitcomposite/clickit_composite.pdf [accessed July 23, 2006.]

sudden drops in fatality rates), and
- Other anomalies that trigger focused follow-up questions.

NTHSA's measurement method also seems to encourage others to replicate its analysis and targeting practices. For example, Alabama decided to focus on belt use among truck drivers, a particularly non-compliant population in its state.

NHTSA supports it diagnostic and replication efforts with extensive sub-national measurement efforts, including experiments measuring before-and-after behavior change and controlled experiments.

Attribution:
The approach NHTSA takes to reporting – reporting outcomes and interim outcomes in individual states and nationally and tracking state uptake of specific laws and practices NHTSA has found to be effective – makes clear what each state contribute to fatality reductions. No explicit calculation of the marginal contribution of each level of government is done.

NHTSA often does small-scale experiments to asses the effect of specific assistance efforts. It helped South Carolina adapt the "Click-It-or-Ticket" media campaign used in North Carolina to work in a state without primary enforcement authority. The replication attempt worked in South Carolina and later in all the southeastern states. Belt usage increased 9 percent in the southeastern states.[100]

To confirm that the effect was attributable to the campaign, NHTSA recruited volunteer states to participate in a controlled, measured experiment. Ten states tested the NHTSA "Click-It-or-Ticket" campaign, four states served as a control group that did nothing, and four states tested programs of their own design. NHTSA-funded observers to measure belt use before and after the campaign using a common measurement methodology. Belt use increased 8.2 percent in full implementation states, 2.7 percent in states using programs of their own design, and 0.5 percent in the control group of states.[101]

Comprehensiveness:
NHTSA has omitted a key outcome indicator: injuries. As the military is discovering with the Iraq war, fatality rates can plummet (because of innovation in quick-response medical teams) while injuries and the cost of post-event care soar. (Gawande) Not knowing this information could lead NHTSA to choose a fatality reduction goal when it should be considering, or at least bringing to the attention of the public for debate, the need to shift some attention to injury reduction. Fortunately, about a dozen states collect non-fatal accident data that NHTSA is able to use to assess whether more attention should be paid to non-fatal injuries.

Implications/ Ideas for EPA	- NHTSA uses a common outcome-focused measure of a single unwanted event for all of its programs and notes key characteristics of the event, using the Haddon matrix, to help it set priorities and find preventable causal factors. EPA might experiment with the feasibility of focusing on a single or a few high priority outcome measures it wants to prevent, such as significant releases to the environment or key non-compliance problems, and try to develop a framework similar to the Haddon matrix to better understand priority problems needing assistance or enforcement attention. - NHTSA sets specific targets for specific subsets of problems, populations, and practices. EPA might benefit by similarly setting specific CA targets for population, problems, or practice areas.

	- NHTSA compares performance across different delivery units to identify those with stronger performance and then "drills down" to understand why. EPA might similarly look at differences in reported levels of understanding, behavior change, and pollution reduced across its various performance units, especially those serving similar populations, to determine if the variation is due only to reporting differences or to different practices. - Although NHTSA does not use state adoption of proven practices such as helmet laws as a national performance measure, it might be a useful national measurement strategy for EPA in areas it chooses to target. Once it has identified effective compliance assistance practices, EPA might track uptake of those practices in other areas to indicate and possibly extrapolate program success for a specific targeted area. - NHTSA integrates measurement of its assistance and (state/local) enforcement efforts, focused on specific outcome targets for specific problems. EPA has begun to do this for its national priorities, but may want to take it the next step to measure assistance and enforcement efforts in an integrated way.

Agency /Program and Relevance to EPA Compliance Assistance	**Internal Revenue Service Taxpayer Service Assessment**
Relevant Performance Goal/Objective from Strategic Plans	IRS Strategic Plan[102] Goal 1 – Improve Taxpayer Service Objectives • Improve service options for the tax paying public • Facilitate participation in the tax system by all sectors of the public • Simplify the tax process Note – the IRS strategic plan has the tag-line "Service + Enforcement = Compliance" The IRS has structured its goals accordingly, with one goal for service (described above), which includes CA, and a separate goal for enforcement.
PART Targets and Measures Related to Compliance Assistance	**PART MEASURES for Internal Revenue Service Taxpayer Service Assessment:[103]** *Outcome Measure:* **Telephone Level of Service** Target: by 2005, 82% of the time taxpayers calling IRS toll-free operations successfully reach a live assister *Output Measures:* **Customer Satisfaction with Telephone Service** Target: by 2006, 94 % of customers responding they are either satisfied or very satisfied with IRS' telephone service (based on surveys of a random group of recent customers) **Customer Satisfaction with Walk-in Service** Target: by 2007, 91% of customers responding they are either satisfied or very satisfied with IRS' walk-in service (based on surveys of a random group of recent customers). **Customer Satisfaction with Correspondence Service**

---

[102] IRS Strategic Plan 2005-2009, http://www.irs.gov/pub/irs-utl/strategic_plan_05-09.pdf

[103] PART Review for Internal Revenue Service Taxpayer Service Assessment,
http://www.whitehouse.gov/omb/expectmore/detail/10002233.2004.html

	Target: by 2008, 69% of customers responding they are either satisfied or very satisfied with IRS' correspondence service (based on surveys of a random group of recent customers). **Tax Law <u>Accuracy</u> for Telephone Service** Target: by 2009, 91.4% of answers provided by IRS telephone assisters on tax law questions are accurate **Tax Law <u>Accuracy</u> for Walk-in Service** Target: by 2009, 80% of answers provided by IRS field assisters are accurate **Accounts <u>Accuracy</u> for Telephone Service** Target: by 2009, 93.5% of answers provided by IRS telephone assisters on taxpayer accounts questions are accurate **<u>Accuracy</u> of Adjustments and Responses to Taxpayer Correspondence** Target: by 2009, 89.1% of adjustments to taxpayer accounts and answers to taxpayer issues provided by correspondence units which are accurate. *Efficiency Measure:* **Customer Contacts Resolved Per Staff Year** Target: by 2009, 7907 taxpayer contacts resolved for IRS customer service programs (phones, internet, walk-in) divided by FTE used
Methodology Agency Uses to Assess Metrics	*Measurement approaches as outlined in strategic plan,*[104] *for those metrics included in the PART:* <u>Level of Service</u> -the relative success rate of taxpayers calling for assistance and seeking services from a Customer Service Representative. Part of the calculation of results for this measure includes the percentage of call attempts made by taxpayers compared to the number of calls answered by IRS. <u>Customer Satisfaction Data</u> - customers' overall level of satisfaction with key services provided by the IRS, obtained through telephone and mail surveys. Data is also captured for IRS by the University of Michigan Business School's National Quality Research Center for the American Customer Satisfaction Index and by Roper Starch Worldwide, a public opinion polling firm. <u>Rate of Accuracy</u> - the percentage of customers receiving accurate responses to their tax law inquiries and account inquiries. IRS uses the data to evaluate the regulatory accuracy of IRS services. IRS intends to add the measure of accuracy for its Tax Assistance Centers to this calculation. *Additional measurement approaches described in the strategic plan, for measures not addressed by PART:* <u>Burden Reduction</u> - measurements of time and out-of-pocket expense taxpayers incur in meeting their tax responsibilities. <u>Rate of Electronic Interactions</u> - measurements of electronic filing and payment participation rates. <u>Timeliness of Responding to Customer Inquiries</u> - measurements of the time taxpayers wait on the telephone when calling IRS about their accounts or inquiring about tax laws when preparing tax returns; the time from account creation to disposition for taxpayers needing account resolution assistance; and, the response time for those taxpayers who communicate electronically with the IRS.
Authors' Assessment of Pros/Cons of Metrics and Measurement Approach	**Feasibility of measuring accurately and credibly:** Self-reported data should be a reliable source for customer satisfaction outcomes. IRS uses some objective, observable measures (e.g., wait times) that should be reliable. It is not clear what methodology IRS uses to develop reliable estimates of CA service rate of accuracy or burden reduction.

---

[104] IRS Strategic Plan 2005-2009, http://www.irs.gov/pub/irs-utl/strategic_plan_05-09.pdf

	Motivational value:
	IRS' measures should be motivating to staff, as well as regulated entities and other CA providers. For example, any indications of declining customer satisfaction should motivate IRS employees to improve performance, while low rates of participation in electronic filing may motivate other service providers to enhance their services.
	Replication value:
	IRS's measures are transferable to other government agencies offering customer service.
	Coherence:
	The metrics are easily understood.
	Diagnostic value:
	The measures provide high diagnostic value. For example a low accuracy rate indicates a clear need to update the CA with correct information.
Implications/ Ideas for EPA	IRS relies on surveys of CA users to assess CA outcomes. IRS does not appear to address "reach," i.e., the extent to which the agency is getting its materials into the hands of those that need compliance assistance. This is consistent with the current EPA approach.
	IRS does break down its accuracy and customer satisfaction measures by type of CA provided, e.g., telephone, walk-in, and correspondence service. This type of approach could be relevant for EPA; for example, it may be worthwhile to develop parallel but separate measures for customer satisfaction with EPA websites, workshops, site visits, etc.
	EPA could consider exploring the use of the American Customer Satisfaction Index to evaluate its CA programs and to tailor its programs to best serve CA users.

Agency /Program and Relevance to EPA Compliance Assistance	**Small Business Administration -- Small Business Development Centers** The SBA seeks to facilitate the creation and success of small businesses. The SBA Office of the National Ombudsman (ONO) assists small businesses by acting as a liaison between small businesses and regulatory agencies such as the EPA.[105] It also directs small business owners toward the compliance assistance offices of various federal regulatory agencies. The SBA's Small Business Development Centers host compliance programs for small business owners.[106]
Relevant Performance Goal/Objective from Strategic Plans	The Small Business Administration 2008-2013 Strategic Plan[107] does not include goals that specifically mention compliance assistance. It implicitly mentions compliance assistance in Section 3.1, which discusses minimizing the regulatory burden on small businesses. ONO acts as a liaison between small businesses and federal agencies. It allows small businesses to bring their complaints to federal agencies and provides small businesses with a website where small business owners can find links to government services and business compliance information and forms. SBA's 2006 Performance & Accountability Report lists a similar goal (Long Term Objective (LTO)

---

[105] National Ombudsman – Overview,
http://web.sba.gov/ombudsman/ONO%20Web%20Show%20V1.1.htm
[106] PART Review for Small Business Development Centers – Small Business Administration,
http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html
[107] U.S. Small Business Administration 2008-2013 Strategic Plan,
http://www.sba.gov/idc/groups/public/documents/sba_homepage/serv_strategic_plan_2006.pdf

	1.2), which aims to use information technology to facilitate interaction between small businesses and the federal government. [108]
PART Targets and Measures Related to Compliance Assistance	**PART Targets and Measures for Small Business Development Centers Assessment:** [109] *Outcome Measures* **Long-term clients counseled.** Target: by 2009, 55,000 long term clients will be counseled per year. **Percent customer satisfaction.** Target: by 2007, 73.0% of all customers will rate their experience with SBDC favorably. *Explanation:* SBA completed the first year of its division-wide impact study for FY 2004 so FY 2004 represents a change from previous years. [110] *Outcome Measures* **Number of clients counseled.** Target: by 2007, 171,550 clients will be counseled per year. *Explanation:* The allocations for nascent entrepreneur and existing business clients are estimates based on a weighted formula. [111] **Number of attendees trained** Target: by 2007, 395,961 attendees will be trained per year. *Explanation:* The allocations for nascent entrepreneurs and existing business clients are estimates based on a weighted formula. [112] **Number of training hours** Target: by 2007, 1,854,155 training hours will be used per year. **Number of counseling hours.** Target: by 2007, 1,014,819 counseling hours will be used per year. *Explanation:* The allocations for nascent entrepreneurs and existing business clients are estimates based on a weighted formula. [113] *Efficiency Measure* **Cost per client served.** *Explanation:* The allocations for nascent entrepreneurs and existing business clients are estimates based on a weighted formula.
Methodology Agency Uses to	Each quarter, SBDCs submit counseling and training data to SBA. They are also required to submit annual economic impact data and semi-annual and annual reports

---

[108] SBA 2006 Performance & Accountability Report,
http://www.sba.gov/aboutsba/budgetsplans/serv_abtsba_budgets_par2006.html
[109] PART Review for Small Business Development Centers – Small Business Administration,
http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html
[110] PART Review for Small Business Development Centers – Small Business Administration,
http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html
[111] PART Review for Small Business Development Centers – Small Business Administration,
http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html
[112] PART Review for Small Business Development Centers – Small Business Administration,
http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html
[113] PART Review for Small Business Development Centers – Small Business Administration,
http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html

Assess Metrics	that chart their progress. Some SBDCs use different data definitions, which complicates SBA's use of the data. [114] Exact methodology for data collection was not described, most likely because SBDCs utilize different data collection methods. As there are over 1,000 SBDCs nationwide, it would not be feasible to list their unique methodologies.
Authors' Assessment of Pros/Cons of Metrics and Measurement Approach	**Feasibility of measuring accurately:** There are 1,100 SBDCs nationwide, which assist over 18 million people. [115] The decentralized system is effective in allowing SBA to assist and interact with small business owners nationwide. However, the system, as it stands, does not lend itself well to accurate data collection. SBDCs use different data definitions, which complicates data analysis and introduces the possibility of error. [116] The weighted formula accounting for the differences between new and existing clients improves the accuracy of the data analysis by accounting for the differing needs of these two types of clients. **Motivational value:** SBDCs have very clear targets to reach: they must create or retain 500,000 jobs by 2007. [117] They also have goals to meet regarding the counseling and training of existing small businesses. These specific, ambitious targets, which are included in the PART review measures, are tracked quarterly and hold SBDCs accountable for their actions and progress made. SBDCs enter into a co-operative agreement with the SBA when they receive funding, and part of the agreement requires that they counsel and train a certain number of clients. If an SBDC fails to meet goals set in the co-operative agreement, it is possible that the SBDC will not be renewed. [118] The measures included in this review, such as number of clients counseled and number of training hours, all relate to the goals that SBDCs have to meet in order to maintain funding, making this measurement approach extremely motivational for SBDCs. **Replication value:** These methods are transferable to other government agencies with regional offices. Because SBA's measurement approach relies on self-reporting by each center, it is less costly than SBA employees conducting a survey of the centers. **Coherence:** Most metrics are easily understood by the public and all clearly reflect SBA and SBDC's goals. After each performance goal listed in SBA's Strategic Plan[119], there is a list of indicators that will be used to evaluate progress toward the goal listed above. Many of those indicators are included in the PART review and are obviously related to the performance goals. **Diagnostic value:**

[114] PART Review for Small Business Development Centers – Small Business Administration, http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html
[115] PART Review for Small Business Development Centers – Small Business Administration, http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html
[116] PART Review for Small Business Development Centers – Small Business Administration, http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html
[117] PART Review for Small Business Development Centers – Small Business Administration, http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html
[118] PART Review for Small Business Development Centers – Small Business Administration, http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html
[119] U.S. Small Business Administration 2008-2013 Strategic Plan, http://www.sba.gov/idc/groups/public/documents/sba_homepage/serv_strategic_plan_2006.pdf

	Measures such as number of counseling hours and number of clients counseled can be helpful in assessing the progress of SBDCs' training and counseling programs, particularly when the weighted formula is taken into account. The measurement approach has only moderate diagnostic value because changes in these measures do not clearly indicate problem areas. A change in the number of counseling hours or clients counseled, for example, could be attributed to a number of factors. **Attribution**: Since SBA funds the SBDCs, it may be possible to attribute benefits to SBA investment and SBDC actions. SBDCs work closely with SBA and non-SBA programs and local economic development partners; SBDCs that are located on a college or university campus receive support from their host institutions. [120] It should be noted that these programs and partners may have an influence on the data gathered, as SBDCs sometimes partner with each other or with other programs to hold workshops. With respect to output measures and one outcome measure (long term clients counseled), benefits are likely attributable to SBA funding and SBDCs' actions. The outcome measure "number of jobs created" may be the result of a number of different factors, including the economy, so it cannot be attributed specifically to SBDCs or SBA. The efficiency measure "Cost per client served" is attributable to SBDCs and their use of SBA funds.
Implications/ Ideas for EPA	The Small Business Administration sets clear, outcome-focused goals for SBDCs, and their progress toward meeting these goals is reviewed regularly. In addition, all the measures included in this review are easily understandable and are easy to relate to the goals outlined in the administration's strategic plan. The effectiveness measure, "Cost per client served" is particularly effective in showing how SBDCs are making use of SBA funds. Implementing a similar measure in all of EPA's compliance assistance programs would highlight the cost-effectiveness of each program and would also make it possible to compare programs' cost-effectiveness.

---

[120] PART Review for Small Business Development Centers – Small Business Administration, http://www.whitehouse.gov/omb/expectmore/detail/10000368.2004.html

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	Description/Objective	Results	Measure and Measurement Methodology	Issues	Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance	Tools Used
California Air Resources Board (ARB) **Project Title:** Air Toxics Compliance Monitoring. **Target Audience:** Chrome plating facilities.	ARB conducted baseline inspection, followed by inspections after three years to compare baseline compliance rates with final compliance rates.	22% of facilities in full compliance during the initial inspection. During final inspection three years later, after CA, 85% of facilities in compliance.	Routine inspections selected through randomized sampling, providing statistically significant compliance rate for chrome platers in region ARB used outcome-based performance measures based on EPA's enforcement and compliance assurance program's National Performance Measures Strategy.	Resource-intensive and not feasible for all cases. Impossible to determine which tools were most helpful.	Small businesses are largely unaware of environmental dangers posed by their work. During course of program, law changed requiring inspections of certain facilities. Facilities inspected and penalized based during baseline study.	1) Integrated approach using permitting, increased inspection, and enforcement plus compliance assistance. 2) Simplifying permit language, handbooks in English and Spanish.
California, San Diego County Department of Health **Target Audience:** Biotechnology and research & development.	San Diego decided to develop an alternative measure using inspection findings to gauge the effect of its compliance assurance work with the biotech and research and development industry.	Tracking this information enabled the County to see the effectiveness of its compliance assistance efforts and quickly detect when a favorable trend line starts to change.	San Diego measured the average number of violations per inspection and the average "top 10" violations per inspection, as well as the total number of violations and top 10 violations. Builds on NH CO RCRA measurement.			
Colorado Department of Public Health and Environment **Project Title:**	Follow up on COMPASS Asphalt pilot. An intervention did not take place during this study. This was a measurement	The report does not go into detail about the size of the effect. It states that identifying trends allowed CDPHE to	Comparison of "traditional compliance inspections" with newer "focused site visit approach" using percent compliance with critical indicators, with	Resource-intensive at front end to develop short survey, but survey format was easy for reviewers to		

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	**Description/Objective**	**Results**	**Measure and Measurement Methodology**	**Issues**	**Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance**	**Tools Used**
Comprehensive Measurement Tool (COMET) project. **Target Audience:** Asphalt plants.	project. The study compares the results of the 2001 and 2002 COMPASS project data according to a set of critical indicators. It also uses a lower cost compliance monitoring approach.	focus compliance assistance on controlling fugitive dust, which led to increasing compliance in the targeted area. Project provided an assessment of each plant's compliance status at the time of inspection, and facilitated identification of compliance trends by grouping questions in related compliance areas together.	information gathered using a more detailed, close-ended questionnaire approach.	understand, even if they were not familiar with the inspection process. Inspection notes from previous years not complete, so could only compare "critical" indicators Fluctuations in compliance may have been due to outside influences. Past data showed increases in compliance even when no CA was provided, so cannot attribute change to CA.		
Colorado Department of Public Health and Environment	Test of value of self-reported measurement.		Colorado developed the Self-Certification and Reporting (SCORE), borrowing from the MA	Survey takes 1 -2 hours for regulated facilities to complete. Many		

[121] Shewmake 2004.

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	**Description/Objective**	**Results**	**Measure and Measurement Methodology**	**Issues**	**Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance**	**Tools Used**
Project Title: Self-Certification and Reporting (SCORE)[121] **Target Audience:** Small Quantity Generators (SQGs)			Environmental Results Program. Facilities fill out a survey.	reported it was a helpful checklist and refresher on hazardous waste regulations.		
Colorado Department of Public Health and Environment **Target Audience:** Large Quantity Generators (LQGs)	To improve or better measure compliance results in the hazardous waste area. In 1999, CO increased its inspection rate to inspect its 150 LQGs at least once every three years because it was troubled by the 70 compliance rate. By 2004, despite the increased inspections and anecdotal evidence of improvement, compliance rate remained about the same, so CO wanted to develop a method that allowed it to detect a different dimension of	In 2000, facilities in violation generally had eight or more violations. In 2004, facilities in violation generally had only four violations.	Colorado traditionally counts a facility non-compliant if it has any violations during the year. The state found that over time, the same number of facilities had at least one violation, but the overall number of violations had dropped.			

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	**Description/Objective**	**Results**	**Measure and Measurement Methodology**	**Issues**	**Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance**	**Tools Used**
	change.					
Connecticut Department of Environmental Protection **Target Audience:** Auto recycling industry; Small Quantity Generators and facilities with USTs.	CT DEP developed a compliance guide for auto recyclers, including a Stormwater P2 model plan.	The average compliance rate for each question improved from the baseline year for all 5 indicator questions between roughly 2 – 15 percent.	Questionnaire for auto recyclers to assess training effectiveness. The success of training was measured by 5 indicator questions (e.g., Do you have a stormwater prevention plan?). For SQG and UST, created (but have not yet used) compliance indicator surveys to create baseline data.	It is unclear whether compliance assistance caused compliance increase.		1) Compliance guide for auto recyclers developed with industry input, including a Stormwater P2 model plan. 2) After guides distributed, hosted 4 training sessions.
Connecticut General Permit **Project Title:** Phase I - Discharge of Minor Tumbling of Cleaning of Parts Wastewater Phase II –Discharge of Minor Printing and Publishing Phase III – audit of General Permit to Limit Potential to Emit.	Assess industry compliance, identify root causes of non-compliance, and develop and employ CA and enforcement strategies to raise compliance rates, identify opportunities for program improvement. PI - Sent information about regulatory requirements to all permit holders. Initiate enforcement	PI - Compliance levels went up dramatically on 4 of 5 parameters, and already high on 5th. PII- Number of printers in system went from 60 to 202. PIII – measured compliance, types of non-compliance,	PI – Mailed to universe of known regulated parties PII – Research to find unregulated parties, then mail to all of them. PIII - Divided state into 5 zones, randomly selected 3. Randomly select 24 facilities per zone. 29 % agree to participate. Pairs selected to minimize travel time and costs (2/day.) DEP sends letter asking facilities to participate voluntarily in confidential audit by private			Tumblers - Mailed material about regulatory obligations paired with enforcement. Advance warning of plan to conduct follow-up monitoring.

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	Description/Objective	Results	Measure and Measurement Methodology	Issues	Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance	Tools Used
	against serious offenders. Two years later, follow up with reminder to submit monitoring data. P II - Mailed to unpermitted printers to get them to register warning of future follow-up. Inspected 30 facilities that did not respond. P III - Project to test third-party confidential measurement approach, not to change compliance levels.	and reasons for non-compliance.	third-party to determine overall compliance levels, type of non-compliance, and root cause for non-compliance.			
EPA Region 1 **Project Title:** EPA New England's Public Works Initiative **Target Audience:** Departments of Public Works (DPWs)	Paired self audits with compliance assistance through trainings and meetings with DPWs.	More than 90% of the DPWs that conducted a self-audit found and corrected environmental compliance issues that they discovered during their self-audits.	The study surveyed 51 of the 216 DPWs that participated in the self-audit in order to evaluate the effectiveness of the initiative. Survey respondents in the Industrial Economics paper were selected randomly, and the sample size is	The Harvard study's interviews are not statistically significant and may include self-selection bias as a result of people who did not return the study authors' calls.	DPWs used compliance assistance either because they felt an ethical obligation to be compliant with environmental regulations or because they wanted to avoid having to pay fines.	Outreach tools, (workshops with APWA and the checklist given to DPWs.)

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	**Description/Objective**	**Results**	**Measure and Measurement Methodology**	**Issues**	**Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance**	**Tools Used**
		More than 90% of the DPWs that participated in the AI reported and corrected the environmental violations they found during their self-audits, and over 80% plan to conduct similar audits in the future. In addition, 75% of respondents are beginning to implement beyond-compliance programs.	statistically significant. Measures used include: percentage of participants correcting environmental violations and participants deciding to undertake beyond-compliance and long-term environmental management activities.			
EPA Region 1 **Project Title:** Colleges & Universities Initiative (the C/U initiative). **Target Audience:** Colleges & Universities.	To improve environmental management and performance, as well as to promote environmental awareness on college and university campuses. Phase 1 involved providing colleges and universities with basic	Over 75% of schools took short-term action to correct problems found during their self-audits, and over 95% of schools made long-term changes to their environmental management strategies as a result of self-audits.	Measured compliance assistance through surveys and interviews, which provided quantitative and qualitative data insights. Survey of colleges and universities in the C/U initiative, using a stratified random sample. Interviews of both initiative participants and non-		The materials that sought to publicize and clarify regulations (such as workshops, common violations list, and regulatory interpretations) were viewed as the most useful and effective in assuring compliance assistance. The biggest incentive	

83

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES					
Government Agency, Program, Target Audience	Description/Objective	Results	Measure and Measurement Methodology	Issues	Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance	Tools Used
	compliance information and used targeted enforcement efforts to stress the importance of compliance. The Audit Initiative (AI) and the Environmental Management System (EMS) Initiative began during Phase 2. For AI, schools conducted self-audits and corrected any violations they found. The EMS initiative encouraged schools to implement EMSs. Phase 3 involved beyond-compliance activities that helped to make campuses more sustainable.	Respondents wanted: more convenient workshop locations, a more user-friendly website, and tailored outreach materials.	participants. Using interviews to determine what led respondents to not participate in the C/U Initiative allowed EPA to learn from the participants answers, rather than trying to predict their answers.

Measures related to: participation in audit initiative, the program's impact on environmental management, and long-term environmental management changes. | | was a lower inspection priority and decreased fines for colleges and universities. Some institutions also said that they participated for ethical reasons. | |
| EPA Region 2

Project Title:
Healthcare Compliance Initiative: | The hospitals had limited resources for carrying out training activities, lacked an environmental compliance chain of | Workshops, trainings, and presentations were the most useful according to survey respondents (28%), | Amount by which pollution is reduced, the number of violation and violations corrected, the number of enforcement actions, and understanding of | The brief report does not give information about how respondents were surveyed or how the amount of | Most regulated parties decided to participate in a voluntary audit or self-disclosure in order to avoid enforcement action | Integrated strategy including:

1) Compliance assistance |

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	Description/Objective	Results	Measure and Measurement Methodology	Issues	Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance	Tools Used
Environmental Results **Target Audience:** Hospitals.	command, were unaware of environmental regulatory requirements, and used hazardous materials even when adequate substitutes were available. EPA's Region 2 decided to use the combination of tools listed above to bring hospitals into compliance, to make compliance a priority, and to extend the compliance improvements to all facilities, not just those receiving inspections.	followed closely by websites provided by Region 2 (27%), outreach materials (24%), and phone calls or emails to EPA (21%). Over 40,000 lbs/year of hazardous waste is managed appropriately, thanks to the compliance assistance initiative undertaken by Region 2. There were significant gains in environmental management of oil, lead-based paint, and CFCs.	regulations.	pollution reduced was estimated.	(41%). Others chose to do so in order to receive reduced penalties (32%) and due to a feeling of moral responsibility (27%).	2) Compliance incentives 3) Compliance monitoring 4) Enforcement 5) Pollution prevention 6) Environmental management systems
Measuring the Performance of E-Government (Report by Genie N.L. Stowers)		Pop-up surveys have higher response rates than opt-in surveys.	E-government performance measured using telephone surveys and online pop-up surveys or clickable "opt-in" surveys.	Many agencies cannot use full evaluation possibilities due to privacy policies.		

85

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	**Description/Objective**	**Results**	**Measure and Measurement Methodology**	**Issues**	**Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance**	**Tools Used**
Target Audience: Government agencies.			Web tracking to determine number of first-time visitors & repeat visitors, the percent of visitors that click on a particular link, etc.			
King County Department of Natural Resources and Parks **Project Title:** King County Environmental Behavior Index (EBI) **Target Audience:** King County residents.	EBI is a survey of households in King County designed to gather information on their behaviors related to yard care, recycling, disposal, and environmentally friendly purchasing. There are 30 indicators in total. Each is evaluated on a color scale that includes: bright green, light green, yellow, brown, grey (ranging from "do the desired behavior all or most of the time" to "don't know about the behavior or what their household is doing"), and white ("does not apply").	Five of the 29 behaviors that were compared to the 2005 baseline data have changed significantly since then. There was an increase in recycling (65% vs. 60%) which may be attributed to new regulations in Seattle. The other increases were in: condom disposal, disposal of latex or water based paints, stains, sealers; and proper treatment of trees and shrubs for insects/diseases.	Conducted a telephone survey of 1000 King County residents. Random Digit Dialing (RDD) was used to determine which households to call. In order to achieve geographic diversity, calls were distributed among the following areas: Seattle, other incorporated areas in King County, and Unincorporated King County. Researchers weighted the data so that it would more closely reflect the adult population of King County (respondents were more often female, older, and at a higher education and income level than the public).	Baseline data is only 1 year old; further studies are needed to see if current trends are valid or anomalies. Some residents say they practice more environmental behaviors than they do, possibly because they are unaware or want to appear more environmentally friendly. More research with target audiences is needed to assess the level of inaccuracy in self-reported data.	People who are already adopting environmental behaviors some of the time (color code: light green & yellow) are most likely to participate in and benefit from compliance assistance.	Telephone surveys.

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	Description/Objective	Results	Measure and Measurement Methodology	Issues	Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance	Tools Used
Massachusetts Department of Environmental Protection. **Project Title:** Environmental Results Program. **Target Audience:** Printers; Photo processors; Dry cleaners.	MA ERP includes compliance assistance, through workbooks. This study focused on the development of a data base, spreadsheet, and software to automate and simplify ERP analysis and assess environmental outcomes impact of ERP programs.	The tools have been developed and are available for use by other agencies.	Universe identification. Self-reporting and certification of reporting accuracy required by all regulated entities on checklist of compliance obligations and P2 practices. Statistically valid inspections conducted by state to determine error rate of self-reported material.			Sector-specific workbooks.
Massachusetts Department of Agricultural Resources, Pesticide Bureau **Project Title:** Integrated Pest Management **Target Audience: Schools and day care centers.**		There was no correlation between schools that attended the workshops given and schools that were in compliance. Some of the positive comments from the online survey said that the online workbook was very helpful and that the website would most likely make it easier to get facilities to submit	Online surveys, onsite visits and follow-up visits, and on site testing of IPM knowledge used to assess the project's progress. The project also used the number of visits to the project's website and the number of people attending educational sessions to ascertain the program's usefulness. The website contains an IPM plan development tool, with a survey built in that asks respondents about the amount of time it took them	The methods cannot be analyzed statistically, so they cannot be used to conclusively prove or disprove their effectiveness in improving the compliance rate.	Some schools came into compliance because pesticide applicators were not allowed to work in the schools unless the school had complied with the IPM law. Others came into compliance due to fines for violations, or the threat of fines for violations. Many school administrators and day care personnel had not ever heard of an IPM	

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	Description/Objective	Results	Measure and Measurement Methodology	Issues	Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance	Tools Used
		IPM plans.	to create their IPM plan and solicits comments and critiques on the site.		plan.	
Maryland Department of the Environment. **Project Title:** Park Heights Project. **Target Audience:** Auto body and mechanic repair facilities in low-income and minority communities.	The first objective of the project, devising an accurate compliance rate calculation methodology, was successful. The second, improving the compliance rate through compliance assistance, could not be proved to be successful.	Combined inspections and evaluations helped MDE learn from the owners, what constituted effective compliance assistance.	Used statistics to determine how many random samples were needed to prove or refute the study's null hypothesis that compliance assistance improved the compliance rate. Random inspections were carried out with an inspector checklist where inspectors recorded observations on indicators. Those observations were used to determine the baseline compliance rate. Follow up inspections were carried out using the same random sampling methodology, visiting more than 30 locations (as in the baseline inspection).	High turnover among regulated community, resulting in too small a sample to disprove the null hypothesis. Larger sample size needed in high turnover sectors. Inspectors were used to focusing on violations while doing inspections. They had trouble using knowledge indicators section of inspection checklist for baseline inspections in order to find out what owners knew and did not know about the regulations facing them.	First version of workbook too complicated. Workbooks translated into "plain English" the owners were more receptive to the information they contained. Considered questions owners and operators were asking while rewriting workbook. Voluntary self-certification not as effective as anticipated; mandatory self-certification is recommended by report. Owners wanted to know how to avoid penalties.	1) A user-friendly workshop for all interested facilities 2) Voluntary self-audits with self-certification forms 3) Community survey and shared results with regulated entities; site-specific solutions.

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	Description/Objective	Results	Measure and Measurement Methodology	Issues	Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance	Tools Used
New Hampshire Department of Environmental Services. **Project Title:** Compliance Measures Project	This project sought a more cost-effective approach to RCRA inspections. Compliance surveyors handed out written CA information, and observed 10 items. Compliance surveyors alerted inspectors immediately if serious problems seen.	439 facilities were visited in 10 weeks, which was over 10 times the number of facilities in one year that NH inspectors usually reach Inspection results: 1) An increased demand for CA information. 2) The trade association hired a full time specialist to conduct semi-annual on-site compliance evaluations. 3) Supplier sent letters re: proper management & disposal practices.	Modified inspection form: 7 compliance and 3 beyond compliance questions. Administered by trained summer interns in 1 hour. Random sample of regulated universe constructed after data base cleaned. Violation type noted. Emergency postings and training lowest compliance levels.	The data clean up was done before sampling started.	Awareness of regulatory obligation and perhaps perceived inspection effect, even though the inspection was "lite."	
New Jersey Department of Environmental Protection			Measures compliance rate based on inspections done, simultaneously showing inspection rates.			
New Mexico Environment	Project staff created and distributed	Written compliance activities were more	State inspector determined compliance outcomes of	Some compliance assistance	The nature of the violation and the	1) Phone calls,

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	Description/Objective	Results	Measure and Measurement Methodology	Issues	Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance	Tools Used
Department **Project Title:** Class V UIC Compliance Assistance Project. **Target Audience:** Class V septic system sites.	outreach materials designed to inform owners and operators of Class V septic system permit requirements. The project utilized targeted outreach to owners and operators of permitted and non-permitted septic systems to improve compliance.	helpful in ultimately achieving compliance than phone calls. Many owners and operators of permitted sites voluntarily submitted a form from one of the outreach materials they received. Of the 13 sites with evidence of groundwater contamination before the project began, four had stopped discharging contaminants to groundwater within two years.	compliance activities at Class V septic system sites.	activities, such as inspections, meetings, and emails, were not evaluated under the same rubric as the other compliance assistance activities (ex: phone calls, status letters, enforcement actions, etc.) because the results were not comparable among sites. Measuring compliance outcomes for individual compliance assistance actions is useful but difficult, according to the project report, because facilities are continually going	permit holder's economic and/or educational level, may affect compliance. Obligation to comply with the law drove some sites to participate in this project. Unpermitted sites were sent letters from NMED requiring them to apply for a groundwater permit, and 22 of 39 did so after receiving the letter.	2) Enforcement letters, 3) Status letters, 4) Corrective action, 5) Inspection, and 6) Other personal communication such as emails and meetings

90

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	**Description/Objective**	**Results**	**Measure and Measurement Methodology**	**Issues**	**Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance**	**Tools Used**
				in and out of compliance.		
Oregon Department of Environmental Quality (ODEQ) **Report Title:** General Deterrence of Environmental Violation: A Peek into the Mind of the Regulated Public **Target Audience:** Oregon residents; companies regulated by ODEQ.	1) Determining whether and what aspects of enforcement tools such as enforcement and penalties play an important role in deterring companies and individuals from non-compliance. 2) Determining the importance of general deterrence in encouraging compliance.	64% of companies that responded believe that compliance assistance activities improve compliance rates more than enforcement actions. The survey found that regulated parties who had received a penalty from ODEQ were much more likely to support greater compliance assistance than were parties who had not been penalized. (32% vs.15%) 77% of companies say that companies could save money by not complying, but it is likely that they would get	To evaluate individuals' opinions on deterrence, the Oregon Department of Environmental Quality (ODEQ) split the state into three regions and called 100 homes in each region using random digit dialing (RDD). To evaluate companies, ODEQ decided to focus on the most regulated companies in the state. It pulled all company names from all ODEQ databases and randomly selected 450 to be interviewed. All companies were called during business hours, and multiple attempts were made to reach them. Computer-aided telephone interviewing (CATI) was used to administer the questionnaire to respondents and enter data	Because the companies interviewed were taken from a list of companies already present in DEQ databases, the results may overlook companies that are continually non-compliant and successfully evade permits and regulation.	Top Compliance Motivators (Average Rating): 1) Possibility of forced shut-down (6.3) 2) Environmental concern (6.3) 3) Possibility of criminal prosecution (6.2) 4) Concern for reputation (6.0) 5) Community concern (5.9) 6) Pressure from customers (5.8) 7) Actual fines (5.8) 8) Potential fines (5.8) 9) Pressure from insurers (5.6) 10) Pressure from employees (5.5) 11) Withholding of state or federal contracts (4.2)	Surveys

91

			APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES			
Government Agency, Program, Target Audience	**Description/Objective**	**Results**	**Measure and Measurement Methodology**	**Issues**	**Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance**	**Tools Used**
		caught.	into the project database.			
South Carolina Department of Health and Environmental Control; Industrial Ecology Program, School of the Environment, Univ. of South Carolina **Target Audience:** Auto body shops; Trucking companies; Private transportation firms.	Assistance initiative to increase awareness and compliance among 150 regulated parties.	First inspection, 44% of small businesses inspected in compliance Follow-up, 100% of the businesses in compliance. There was a reported increase in awareness, but it was not measured.	Inspector findings were used to assess compliance among a target group of businesses, all of which were inspected before CA and then reinspected.		High turnover rate of businesses results in non-compliance, as new owners were not aware of environmental regulations or CA programs. More stable businesses are more likely to use CA programs.	1) Website, 2) Hotline, 3) Repeated on-site visits; forty-five day window without enforcement action given.
United Kingdom Environment Agency **Report Title:** Greener Business is Good Business 2003 **Target Audience:** Small and medium sized enterprises (SMEs)	The environment agency provides advice for businesses on preventing pollution and has targeted small and medium sized enterprises (SMEs) to help them improve their environmental performance.	Overall improvements in each sector cannot be directly linked to the compliance assistance provided in the individual case studies, but the data showed that serious pollution incidents had decreased by 15% since 2002. Waste	The report provides data about operator performance, pollution emissions to air and water, and serious pollution incidents. Data is also compared across sectors in the categories listed above. Data comes from the Environment Agency, the National Statistics' Annual Business Inquiry, and DTI's Digest of UK Energy Statistics.	Conducting an assessment like this one would be resource-intensive and would require a high level of cooperation within EPA and across numerous organizations.	Surveys conducted by the Environment Agency found that ¾ of SMEs think that they do not have a negative impact on the environment and are not well-versed in environmental legislation. Regulated parties sought compliance	1) Demonstrations and workshops for farmers, 2) Personalized support (visits to individual facilities, businesses, and farms) 3) Working with

APPENDIX B. MEASUREMENT LESSONS FROM OTHER ENVIRONMENTAL AGENCIES						
Government Agency, Program, Target Audience	**Description/Objective**	**Results**	**Measure and Measurement Methodology**	**Issues**	**Factors Affecting Non-Compliance or Raising Interest in Compliance Assistance**	**Tools Used**
		recovery improved by 6%, and operator performance improved; over half of regulated sites received either an A or B.	The Environmental Agency grades businesses on an A-E scale, and gives those grades in this report.		assistance when they were fined by the Environment Agency.	businesses to develop educational materials that meet their needs
Washington Department of Ecology. Hazardous Waste and Toxics Reduction Program Hazardous waste generators (all sizes)	1) State-wide baseline quantitative measure of environmental compliance 2) Relationship between less rigorous than standard enforcement inspections (CEIs) and regulatory compliance? 3) Relationship between regulatory technical assistance visits and regulatory compliance?	"Strongly suggests" that the longer the time period between inspection and compliance, the poorer regulatory compliance at facilities. LQGs had worse baseline compliance than MQGs, and the lowest type of non-compliance for MQGs and LQGs was spills. Assistance visits seem to be as effective as compliance inspections for MQGs and SQGs.	Sample size: 265 generators. The program created three different, compliance indexes ("score") for regulated generators (Y/N, Categorical, Absolute.) It used weighted baseline comparisons and time stratified comparisons (compared 3 lengths of time from inspections 6 years).	Resource/budget constraints limited site inspections, but statistically significant. Developing queries "highly technical task" that required computer services staff and software. Tracking field work and manual entry of data by regular inspection staff proved difficult.	LQGs had a lower compliance rate than MQGs. Generator status proved the strongest predictor of environmental compliance. Authors identified reasons for increased non-compliance from last inspection: changes in facility personnel, management, and/or company ownership; and forgetfulness or laziness on part of long-term staff.	1) Visits with inspections for some 2) Visits with assistance only for some